RANDOM SUBGRAPHS OF THE 2D HAMMING GRAPH: THE SUPERCRITICAL PHASE

REMCO VAN DER HOFSTAD AND MALWINA J. LUCZAK

Abstract. We study random subgraphs of the 2-dimensional Hamming graph $H(2,n)$, which is the Cartesian product of two complete graphs on n vertices. Let p be the edge probability, and write $p = \frac{1+\varepsilon}{2(n-1)}$ for some $\varepsilon \in \mathbb{R}$. In [4, 5], the size of the largest connected component was estimated precisely for a large class of graphs including $H(2,n)$ for $\varepsilon \leq \Lambda V^{-1/3}$, where $\Lambda > 0$ is a constant and $V = n^2$ denotes the number of vertices in $H(2,n)$. Until now, no matching lower bound on the size in the supercritical regime has been obtained.

In this paper we prove that, when $\varepsilon \gg (\log V)^{1/3} V^{-1/3}$, then the largest connected component has size close to $2\varepsilon V$ with high probability. We thus obtain a law of large numbers for the largest connected component size, and show that the corresponding values of p are supercritical. Barring the factor $(\log V)^{1/3}$, this identifies the size of the largest connected component all the way down to the critical p window.

1. Introduction

We study random subgraphs of the 2-dimensional Hamming graph $H(2,n)$. The d-dimensional Hamming graph is a graph on $V = n^d$ vertices, each corresponding to one of the n^d distinct d-vectors $\mathbf{v} = (v_1, \ldots, v_d) \in \{1, \ldots, n\}^d$. A pair of vertices are connected by an edge if and only if these vertices differ in precisely one coordinate. (See for example [9] for more information on the properties of Hamming graphs.) The 1-dimensional Hamming graph $H(1,n)$ is the complete graph; for $d \geq 2$, the graph $H(d,n)$ is the Cartesian product of d complete graphs on n vertices. In particular, it is transitive and the degree of each vertex is $\Omega = d(n-1)$.

We write $\mathbb{P}_p$ for the probability law of the random subgraph of $\mathbb{G}$ resulting when each edge is *occupied* (or *present*) with probability p, and *vacant* (or *absent*) with probability $1-p$, independently of all the other edges. We write $\mathbb{E}_p$ for the expectation with respect to $\mathbb{P}_p$. Also, Var_p will denote the variance under $\mathbb{P}_p$.

Throughout we work with the 2-dimensional Hamming graph $H(2,n)$ unless explicitly stated otherwise, and we shall assume that $p = \frac{1+\varepsilon}{2(n-1)} = \frac{1+\varepsilon}{\Omega}$, where $\varepsilon = \varepsilon(n) \in (0,1)$ tends to 0 in a certain way to be specified below. Our goal is to study properties of random subgraphs of $H(2,n)$ under $\mathbb{P}_p$.

Random subgraphs of finite tori with various edge sets were studied in quite some generality in [4, 5], and we now highlight the key results of these papers. Some of the theorems in [4, 5] apply to a general finite transitive graph, which in what follows will be denoted by $\mathbb{G}$. We also denote the number of vertices or volume of $\mathbb{G}$ by $V = |\mathbb{G}|$ and the vertex degree by Ω. Given a vertex $\mathbf{v}$ of $\mathbb{G}$, we shall write $C(\mathbf{v})$ for the *connected component* or *cluster* containing $\mathbf{v}$, and $|C(\mathbf{v})|$ for the number of vertices in $C(\mathbf{v})$. Further, we let $\chi(p)$ be the expected size of the cluster containing $\mathbf{v}$,

Date: 27 December 2007 (December 15, 2008).
2000 *Mathematics Subject Classification.* 05C80.
Key words and phrases. random graphs, percolation, phase transition, scaling window.

that is

$$\chi(p) = \mathbb{E}_p[|C(\mathbf{v})|]. \tag{1.1}$$

(Note that, by transitivity, this is independent of the choice of $\mathbf{v}$.) Then in [4, 5] the *critical threshold* $p_c = p_c(\mathbb{G}, \lambda)$ of a finite transitive graph $\mathbb{G}$ is defined to be the unique solution to the equation

$$\chi(p_c) = \lambda V^{1/3}, \tag{1.2}$$

where $\lambda > 0$ is a sufficiently small constant. (See [5] for details concerning the precise constraints on the size of λ.)

In [4], cluster sizes were investigated for graphs $\mathbb{G}$ satisfying the so-called triangle condition. In [5], the triangle condition was established for certain types of graphs $\mathbb{G}$, including the Hamming graph $H(d, n)$ of a general dimension $d \geq 1$. We shall now describe these results briefly in order to set up our own scene.

Let $\mathcal{C}_{\max}$ denote a cluster of maximum size, where we may pick any such cluster if it is not unique. Then $|\mathcal{C}_{\max}|$ is the maximum cluster size, that is

$$|\mathcal{C}_{\max}| = \max\{|C(\mathbf{v})| : \mathbf{v} \in \mathbb{G}\}. \tag{1.3}$$

The main theorems in [4] concern the scaling of $\chi(p)$ and bounds on $|\mathcal{C}_{\max}|$ in graphs $\mathbb{G}$ satisfying the triangle condition as $|\mathbb{G}| = V \to \infty$. Specifically, it is shown in [4, 5] that, if p_c is as in (1.2) and

$$p = p_c + \frac{\varepsilon}{\Omega}, \tag{1.4}$$

then, for all ε such that $\varepsilon V^{1/3} \to -\infty$, asymptotically the expected cluster size $\chi(p)$ satisfies

$$\chi(p) = \frac{1 + O(\Omega^{-1}) + O(V^{-1/3})}{|\varepsilon|}. \tag{1.5}$$

With regard to the maximum cluster size, for all $\omega \geq 1$, as $V \to \infty$,

$$\mathbb{P}_p\Big(\frac{\chi^2(p)}{3600\omega} \leq |\mathcal{C}_{\max}| \leq 2\chi^2(p)\log(V/\chi^3(p))\Big) \geq \big(1 + \frac{36\chi^3(p)}{\omega V}\Big)^{-1} - \sqrt{e}[2\log(V/\chi^3(p))]^{-3/2}. \tag{1.6}$$

The above describes the behaviour of the mean and maximum cluster sizes for *subcritical* p values, which are p values satisfying $\varepsilon V^{1/3} \to -\infty$; in particular, the bounds apply to $H(2, n)$.

For a constant $\Lambda > 0$, the *critical window* is defined as the interval of all $p = p_c + \frac{\varepsilon}{\Omega}$ such that $|\varepsilon| \leq \Lambda V^{-1/3}$. Theorem 1.3 in [4] shows that, for some constant $b = b(\Lambda)$, the maximum cluster size inside the critical window satisfies

$$\mathbb{P}_p\Big(\omega^{-1}V^{2/3} \leq |\mathcal{C}_{\max}| \leq \omega V^{2/3}\Big) \geq 1 - \frac{b}{\omega}. \tag{1.7}$$

The corresponding results in [4, 5] are significantly weaker in the case $p = p_c + \frac{\varepsilon}{\Omega}$ where $\varepsilon^3 V \to \infty$ (that is, when p is *above the critical window* or *supercritical*). In particular, only upper bounds on the maximum cluster size are established therein. More precisely, it is proved in [4] that, for all $\omega \geq 1$,

$$\mathbb{P}_p\Big(|\mathcal{C}_{\max}| \geq \omega(V^{2/3} + \varepsilon V)\Big) \leq \frac{21}{\omega}. \tag{1.8}$$

The problem with this result is that it does not imply that p_c as defined in (1.2) actually *is* the critical value, and thus that $p = p_c + \frac{\varepsilon}{\Omega}$ with $\varepsilon^3 V \to \infty$ really is above the critical window. Indeed, to prove that this is the case, one additionally needs a *lower bound* on the maximum connected component size. No such results are established in [4, 5], and we expect that the geometry of the graphs under consideration plays a crucial role in lower bounding the largest cluster size.

The aim of this paper is to establish the asymptotics of the maximum supercritical cluster for the 2-dimensional Hamming graph $H(2,n)$. Throughout our proofs we shall use the phrase "with high probability" (abbreviated as "**whp**") to mean "with probability tending to 1 as $V \to \infty$". Also, "with very high probability" (abbreviated as "**wvhp**") will mean "with probability at least $1 - O(V^{-3})$ as $V \to \infty$". All unspecified limits are as $V \to \infty$. Given an event E, $I[E]$ will denote the indicator of E. We write $\mathbb{P}(\cdot)$ for a generic probability measure (for instance, the probability measure corresponding to a sequence of i.i.d. binomial random variables), which may vary from situation to situation. We use the O_{p} and o_{p} notations in the standard way (see e.g. Janson, Łuczak and Ruciński [18]). For example, if (X_n) is a sequence of random variables, then $X_n = O_{\mathrm{p}}(1)$ means "X_n is bounded in probability" and $X_n = o_{\mathrm{p}}(1)$ means that X_n converges to zero in probability as $n \to \infty$. We shall also use the asymptotic $o(), O(), \Omega(), \Theta()$ notations (without the subscript "p") in the standard way, and again referring to the regime where $V \to \infty$. We write $f(V) \gg g(V)$ (resp. $f(V) \ll g(V)$) when $g(V) = o(f(V))$ (resp. $f(V) = o(g(V))$) as $V \to \infty$. Throughout, the symbol "$\sim$" refers to, often heuristic, estimates of the *leading order* as $V \to \infty$, with unspecified constants and thus uncontrolled error terms. Finally, we denote by C a generic (unspecified) positive constant, which may change from line to line. We shall interchange this use of C with the $O()$ notation.

1.1. **The model.** We consider the Hamming graph $H(d,n)$, and take the edge probability $p = (1+\varepsilon)/\Omega$. We first argue that this agrees asymptotically with the choice of p in (1.4). Let us note that [4, Theorem 1.5] establishes that, for a graph $\mathbb{G}$ satisfying the triangle condition,

$$1 - \chi(p_c)^{-1} \le \Omega p_c \le 1 - \chi(p_c)^{-1} + O(\Omega^{-1}). \tag{1.9}$$

When $\mathbb{G} = H(d,n)$, then $\Omega = d(n-1)$ and $\chi(p_c) = \lambda V^{1/3} = \lambda n^{d/3}$. Therefore, if $\varepsilon = \Theta(V^{-1/3})$, then

$$p = \frac{1+\varepsilon}{\Omega} = p_c + \frac{\varepsilon}{\Omega}(1 + O(1)), \tag{1.10}$$

while for p outside the critical window,

$$p = \frac{1+\varepsilon}{\Omega} = p_c + \frac{\varepsilon}{\Omega}(1 + o(1)). \tag{1.11}$$

Since in the case $d = 2$, we have that $\Omega^{-2} = o(\varepsilon/\Omega)$ for $\varepsilon \gg V^{-1/3}$, the critical value defined in [4, 5] agrees asymptotically to leading order with the value $1/d(n-1) = 1/\Omega$. In particular, $p = 1/d(n-1)$ is inside the critical window of [4, 5]. This shows that we are working in the correct range of p values. For $d \ge 3$, (1.10)–(1.11) may not necessarily be valid, and we shall discuss this issue in more detail in Section 1.2.

From now on, we concentrate on the supercritical case, that is $\varepsilon \gg V^{-1/3} = n^{-d/3}$. Our main result is the following:

Theorem 1.1 (The supercritical phase for $H(2,n)$)**.** *Consider the 2-dimensional Hamming graph $H(2,n)$. Let $p = p_c + \frac{\varepsilon}{\Omega}$ and let $V^{-1/3}(\log V)^{1/3} \ll \varepsilon \ll 1$. Then*

$$|\mathcal{C}_{\max}| = 2\varepsilon n^2(1 + o_{\mathrm{p}}(1)). \tag{1.12}$$

Theorem 1.1 shows that, when $n^{-2/3}(\log n)^{1/3} \ll \varepsilon \ll 1$, the largest connected component satisfies a law of large numbers. Barring the factor $(\log V)^{1/3}$ in the lower bound on ε, Theorem 1.1 identifies the asymptotic size of the largest cluster all the way down to the critical threshold. Therefore, our result demonstrates that $p_c = \frac{1}{2(n-1)}$ really *is* the critical value for random subgraphs of the 2-dimensional Hamming graph. We believe that our proof can be adapted to deal with the case

where $\varepsilon > 0$ is *fixed*. Here, the corresponding statement would be that $|\mathcal{C}_{\max}| \sim \zeta_{1+\varepsilon} V$ **whp**, where ζ_λ is the survival probability of a Poisson branching process with mean offspring λ. Since the proof of Theorem 1.1 is the most challenging for ε as close as possible to the critical window $V^{-1/3}$, we choose not to consider the case of constant ε in this paper.

Before giving a proof of Theorem 1.1, we discuss its statement in more detail in Section 1.2 below. Therein we also include some conjectures concerning Hamming graphs of a general dimension d.

1.2. **Discussion and heuristics.** We first sketch an intuitive picture justifying the definition of p_c from [4, 5] given in (1.2). This picture relies on a branching process approximation for $p < p_c$.

We expect random clusters in our model to exhibit behaviour similar to that of a subcritical branching process. Therefore, from the theory of branching processes, if $p = p_c + \frac{\varepsilon}{\Omega}$ is just below the critical point (for instance, if $\varepsilon < 0$), then we should have (e.g., from the Otter-Dwass formula, see Lemma 3.4 below)

$$\mathbb{P}_p(|C(\mathbf{v})| \geq k) \sim \frac{1}{\sqrt{k}} e^{-\frac{1}{2}k\varepsilon^2(1+o(1))}, \tag{1.13}$$

which in turn implies that

$$\chi(p) = \mathbb{E}_p[|C(\mathbf{v})|] \sim \int_0^\infty x^{-1/2} e^{-\frac{1}{2}x\varepsilon^2(1+o(1))} dx \sim \int_0^{\varepsilon^{-2}} x^{-1/2} dx \sim \varepsilon^{-1}. \tag{1.14}$$

Thus, in fact,

$$\mathbb{P}_p(|C(\mathbf{v})| \geq k) \sim \frac{1}{\sqrt{k}} e^{-\frac{k}{\chi^2(p)}\Omega(1)}, \tag{1.15}$$

and hence for subcritical p (possibly up to logarithmic corrections)

$$|\mathcal{C}_{\max}| \sim \chi(p)^2 \quad \textbf{whp}. \tag{1.16}$$

On the other hand, in the case $p > p_c$ there should be a connected component dominating all the others in size. One way to express this intuitive statement is to impose that

$$\chi(p) = \mathbb{E}_p[|C(\mathbf{v})|] \sim \mathbb{E}_p\big[|C(\mathbf{v})| I[\mathbf{v} \in \mathcal{C}_{\max}]\big] = \frac{1}{V}\mathbb{E}_p[|\mathcal{C}_{\max}|^2]. \tag{1.17}$$

Naturally, the meaning of formula (1.17) is, in essence, that the main contribution to the expected size of a cluster of any particular vertex $\mathbf{v}$ is from those configurations where this vertex lies in the largest component.

Note that (1.17) could be taken as a *defining property* of supercritical behaviour. Then the *critical window* can be defined as the interval of p values where the subcritical and supercritical pictures coincide. In other words, if p lies within the critical window then both (1.16) and (1.17) should be satisfied.

Assume further that a sufficient amount of the concentration of measure exhibited by $|\mathcal{C}_{\max}|$ in the subcritical regime (as implied by (1.16)) carries through to the critical window, so that $\mathbb{E}_p[|\mathcal{C}_{\max}|^2] \sim \chi(p)^4$. It then follows that, for p inside the critical window,

$$\chi(p) \sim \frac{1}{V}\mathbb{E}_p[|\mathcal{C}_{\max}|^2] \sim \frac{1}{V}\chi(p)^4; \tag{1.18}$$

and hence, inside the critical window, we are led to

$$\chi(p) \sim V^{1/3}. \tag{1.19}$$

This provides a rationale for the definition (1.2) of the critical threshold p_c. In conclusion, the above heuristic demonstrates that branching process approximations in the subcritical regime and

the domination of the expected cluster size by the maximum cluster size in the supercritical regime together imply that (1.2) is the "right" definition for p_c.

At this point, we emphasise that subcritical branching process approximations are only likely to be valid for a random graph that is sufficiently mean-field in character, in the sense that its geometry is of little significance for the structure of its random subgraphs. This is the case for sufficiently high-dimensional random graphs, but cannot be expected to hold for low-dimensional random graphs, as indicated in [7, 8]. For random subgraphs of the torus with nearest-neighbour bonds in a sufficiently high (but constant) dimension, as well as for the torus with sufficiently spread-out bonds in dimensions greater than 6, it is shown in [12] that the largest critical connected component is of order $V^{2/3}$, with logarithmic corrections in the lower bound. Accordingly, assuming universality in high-dimensional finite-range percolation, one can expect classical random graph asymptotics at criticality to be valid for random subgraphs of the torus when $d > 6$ for general choices of finite-range edges. On the other hand, the results of [7, 8] suggest that random graph asymptotics at the phase transition threshold are *not* valid for random subgraphs of the d-dimensional torus when $d < 6$.

We close this section with a few comments and conjectures. The present paper verifies the location of the critical window found in [4, 5] up to a factor $(\log V)^{1/3}$. The main barrier to overcoming this separation with our approach is the fact that we require concentration of measure for the number of vertices with either the first or second coordinate fixed in order for our estimates to be sufficiently precise; this concentration property fails when ε is too small. Similar issues cause problems with extensions of our approach to $H(d,n)$ for $d > 2$, although we believe that it could handle the $d = 3$ case. Let us mention at this point that the $(\log V)^{1/3}$ separation has since been removed by Nachmias [22], using non-backtracking random walks. He also manages to nail down the critical window from above in $H(3,n)$, although he does not establish laws of large numbers for the giant component for $H(2,n)$ or $H(3,n)$.

In the present paper we investigate the scaling of the largest connected component in supercritical percolation on the Hamming graph $H(2,n)$. Many random graph models are well known to satisfy what is sometimes referred to as a *discrete duality principle* (see for instance [1, Section 10.5]). This is the principle that the size of the second largest supercritical component is asymptotically close in distribution to the size of the largest subcritical component for an appropriate choice of subcritical edge probability. This notion of duality is closely related to the duality exhibited by branching processes [2, 10, 11, 19], or [1, Section 10.4]. We expect the Hamming graph $H(2,n)$ to follow the discrete duality principle. More precisely, we expect that if we were to remove the largest connected component when $p = p_c + \varepsilon/\Omega$ with $\varepsilon \gg V^{-1/3}$, then the resulting connected components would be like those of the Hamming graph with $p = p_c - \varepsilon/\Omega$. In particular, letting $|\mathcal{C}_{(2)}|$ be the size of the second largest component, we conjecture that

$$|\mathcal{C}_{(2)}| = 2\varepsilon^{-2}\log(\varepsilon^3 V)(1 + o_{\mathrm{p}}(1)). \tag{1.20}$$

For the Hamming graph $H(d,n)$ of an arbitrary dimension d, we conjecture that critical p values are of the form

$$p = \sum_{i=1}^{\lceil d/3 \rceil} a_i n^{-i} + \frac{\mu}{n^{1+d/3}}, \tag{1.21}$$

where λ is an arbitrary constant, and the coefficients $a_i = a_i(d)$ are independent of n. Note that $p_c = 1/\Omega = 1/d(n-1)$ corresponds to $a_i = a_i(d) = 1/d$ for all $i \geq 1$, while $p_c = 1/(\Omega - 1) = 1/(d(n-1) - 1)$, where $d(n-1) - 1$ is the *forward branching ratio of* $H(d,n)$, corresponds to $a_i = a_i(d) = (d+1)^i/d^{i+1}$ for all $i \geq 1$. We believe that, when d is sufficiently

large, there exists an i such that $a_i(d) \neq 1/d$ and $a_i(d) \neq (d+1)^i/d^{i+1}$. In particular, if this is indeed true, then, for $\varepsilon = \Theta(V^{-1/3})$ and d sufficiently large, the edge probability $p = p_c + \frac{\varepsilon}{\Omega}$ is *not* the same as $p = \frac{1+\varepsilon}{\Omega}$ or $p = \frac{1+\varepsilon}{\Omega-1}$. For $d = 2$, however, these choices *do* asymptotically agree, as explained in (1.9)–(1.11).

To explain why we believe (1.21) to hold, we note that [4, 5] indeed gives that the critical window consists of p values given by $p = p_c + \mu V^{-1/3}/\Omega$, that is $p = p_c + \mu n^{-1-d/3}$ on $H(d,n)$. Thus, (1.21) follows for *all* μ, as long as it holds for one particular value of p inside the critical window, for example, for $p = p_c$ defined in (1.2), for any λ, for d fixed and $n \to \infty$. Such asymptotic expansions of critical values in terms of the vertex degree have been established for the n-cube and for nearest-neighbour percolation on $\mathbb{Z}^d$ in [13, 14]. These expansions arise since the value p_c satisfies an implicit equation in terms of certain "Feynman diagrams" occurring in the lace expansion analysis, and these diagrams can be proved to obey asymptotic expansions that in turn imply that p_c has an asymptotic expansion. We expect that this part of the analysis in [13, 14] can be extended to Hamming graphs, and will allow one to compute the numerical values of $a_i(d)$. The proof of this conjecture would enable an extension to random subgraphs of $H(d,n)$ of the phase transition description available for the classical Erdős-Rényi random graph.

For p inside the critical window, $|\mathcal{C}_{\max}|$ is of the order $V^{2/3} = n^{2d/3}$, as proved in [4, 5]. Below the critical window, we expect that the average cluster size satisfies $\chi(p) \sim [\Omega(p_c - p)]^{-1}$, while the maximum cluster size satisfies

$$|\mathcal{C}_{\max}| \sim 2\chi(p)^2 \log\left(V/\chi(p)^3\right) \quad \textbf{whp}. \tag{1.22}$$

Note that [4] establishes in full only the upper bound part of (1.22), the corresponding best lower bound therein being $|\mathcal{C}_{\max}| \geq \chi(p)^2/(3600\omega)$ **whp** for ω large (cf. (1.6)). (It is the upper bound, however, that is relevant for locating the phase transition window.) We anticipate that above the critical window

$$|\mathcal{C}_{\max}| \sim 2\varepsilon V \quad \textbf{whp}, \tag{1.23}$$

where $\varepsilon = \Omega(p - p_c) \gg V^{-1/3}$. Establishing the validity of the asymptotics in (1.23) in full generality would strengthen Theorem 1.1 to all $d \geq 1$ and all p above the critical window.

2. Overview of the proof of Theorem 1.1

This section contains an extensive overview of the proof of our main result, breaking it down into a number of key propositions and lemmas. We start by describing the general philosophy of the proof.

From now on, we shall assume that $p = p_c + \varepsilon/\Omega$, where $\varepsilon \geq 0$. As in [4], the proof will be centered on the investigation of the random variables

$$Z_{\geq k} = \sum_{\mathbf{v} \in H(2,n)} I[|C(\mathbf{v})| \geq k], \tag{2.1}$$

the number of vertices in clusters of size at least k, for appropriate values of k. In terms of these random variables, we have that $|\mathcal{C}_{\max}| \geq k$ holds if and only if $Z_{\geq k} \geq 1$. By proving sufficient concentration of measure for $Z_{\geq k}$, we are able to prove bounds on $|\mathcal{C}_{\max}|$. The whole proof revolves around finding the right scales of k to which we can apply our arguments.

Specifically, we need two *different* scales. The first scale is the smallest possible scale k for which $\mathbb{P}_p(|C(\mathbf{v})| \geq k)$ is very close to 2ε. If indeed the duality principle holds (see the discussion above

(1.20)), then, by (1.15), we expect that

$$\mathbb{P}_p(|C(\mathbf{v})| \geq k) = \mathbb{P}_p(|C(\mathbf{v})| \geq k, \mathbf{v} \in \mathcal{C}_{\max}) + \mathbb{P}_p(|C(\mathbf{v})| \geq k, \mathbf{v} \notin \mathcal{C}_{\max}) \sim 2\varepsilon + \frac{1}{\sqrt{k}} e^{-k\varepsilon^2/2}. \tag{2.2}$$

As a result, as soon as $k \gg \varepsilon^{-2}$, we are led to

$$\mathbb{P}_p(|C(\mathbf{v})| \geq k) \sim 2\varepsilon, \tag{2.3}$$

so that also $\mathbb{E}_p[Z_{\geq k}] = V\mathbb{P}_p(|C(\mathbf{v})| \geq k) \sim 2\varepsilon V$. Equation (2.3) follows from Proposition 2.1 below. Assuming sufficient concentration of measure for $Z_{\geq k}$, we then obtain that $Z_{\geq k} \sim 2\varepsilon V$ **whp**, and, since $|\mathcal{C}_{\max}| \leq Z_{\geq k}$ for every k for which $Z_{\geq k} \geq 1$, we obtain the required upper bound on $|\mathcal{C}_{\max}|$. Concentration estimates on $Z_{\geq k}$ are stated in Proposition 2.2 below.

The lower bound on $|\mathcal{C}_{\max}|$ is slightly more involved. Here we need to find the *largest* possible k for which we can prove that $Z_{\geq k}$ is concentrated around its mean $\mathbb{E}_p[Z_{\geq k}] \sim 2\varepsilon V$. To achieve this, we perform a so-called *two-round exposure*. We first take $p_- < p$ such that $p_- = p + o(\varepsilon/\Omega)$, and compare clusters in percolation with parameter p_- to suitable lower-bounding branching processes. Note that such comparisons can only be applied when $k \ll \varepsilon V$, so these bounds are rather "weak". Subsequently, we "sprinkle" extra edges, so that the distribution of the final configuration is that of percolation with parameter p. We prove that all the large connected components in the p_--configuration are, in fact, **whp**, joined all together by the sprinkled edges. We now explain the steps in this argument in more detail.

Since $p_- < p$ and satisfies $p_- = p + o(\varepsilon/\Omega)$, all the concentration results for $Z_{\geq k}$ hold also for $Z'_{\geq k}$, the number of vertices in connected components of size at least k in the p_-−percolation configuration. Furthermore, again using the fact that $p_- = p + o(\varepsilon/\Omega)$, we have that $\mathbb{P}_{p_-}(|C(\mathbf{v})| \geq k) \sim 2\varepsilon$, so that, by our concentration estimates, $Z'_{\geq k} \sim 2\varepsilon V$ **whp** for all $k \ll \varepsilon V$. This establishes the necessary "weak" bounds on connected components of size at least $k \ll \varepsilon V$.

The p-configuration can be coupled to the p_--configuration as follows. Let $\eta > 0$ be given by $p_- + (1 - p_-)\eta/\Omega = p$. Then, make each p_-−vacant edge occupied with probability η/Ω, independently of all other vacant edges. We show that, for appropriate choices of η (and thus p_-) and $k \ll \varepsilon V$, the sprinkling procedure **whp** connects *all* p_--clusters of size at least k into one. It follows that $|\mathcal{C}_{\max}| \geq Z'_{\geq k} \sim 2\varepsilon V$ **whp**, establishing the lower bound. This part of the proof makes crucial use of the fact that big components turn out to be quite "dense", in the sense that they contain many elements along most coordinate lines; details can be found in Proposition 2.4 below.

As explained above, the entire analysis revolves around a delicate choice of the two different scales. We now present our precise results, formulated in Propositions 2.1, 2.2 and 2.4 below. We then use these propositions to complete our proof of Theorem 1.1.

Proposition 2.1 (The cluster tail)**.** *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Let $V^{-1/3} \ll \varepsilon \ll 1$ as $V \to \infty$. Then, for every η such that $\varepsilon^{-2} \ll \eta V \ll \varepsilon V$,*

$$\mathbb{P}_p(|C(\mathbf{v})| \geq \eta V) = 2\varepsilon(1 + o(1)). \tag{2.4}$$

Proposition 2.1 consists of two parts, corresponding to the upper and lower bounds. These are re-stated separately in Section 4.1 as Lemmas 4.2 and 4.3, and proved in Sections 4.1 and 4.3 respectively.

The following proposition shows concentration of measure for $Z_{\geq k}$ for an appropriately chosen $k \gg \varepsilon^{-2}$.

Proposition 2.2 (Concentration of the number of vertices in large components of certain sizes)**.** *Set $p = p_c + \frac{\varepsilon}{\Omega}$ and let $V^{-1/3}(\log V)^{1/3} \ll \varepsilon \ll 1$. Then there exists ε_0 satisfying $\varepsilon_0 \ll \varepsilon$ such that,*

for every $\delta > 0$,

$$\mathbb{P}_p\Big(|Z_{\geq\varepsilon_0^{-2}} - \mathbb{E}_p[Z_{\geq\varepsilon_0^{-2}}]| \geq \delta\varepsilon V\Big) = o(1). \tag{2.5}$$

The proof of Proposition 2.2 makes use of a somewhat delicate second moment argument. We start by upper bounding the variance of $Z_{\geq N}$ and $Z_{\geq 2N} - Z_{\geq N}$ for specific values of N. These bounds are then combined to prove that $Z_{\geq\varepsilon_0^{-2}}$ is concentrated around its mean, provided that $V^{-1/3}(\log V)^{1/3} \ll \varepsilon \ll 1$. The proof can be found in Section 5.2, where we also show that a possible choice for ε_0 is $\varepsilon_0 = V^{-1/3}$, which indeed satisfies $\varepsilon_0^{-2} = V^{2/3} \gg \varepsilon^{-2}$, since $\varepsilon \gg V^{-1/3}$. For the remainder of this section, we only need to know that a suitable ε_0 does indeed exist; its precise value is irrelevant.

Armed with Propositions 2.1 and 2.2, we now prove the upper bound on $|\mathcal{C}_{\max}|$ in Theorem 1.1.

Proof of the upper bound part of Theorem 1.1. We shall show that **whp**, $|\mathcal{C}_{\max}| \leq 2\varepsilon V(1 + o(1))$. Choose ε_0 as in Proposition 2.2. By Proposition 2.2, the random variable $Z_{\geq\varepsilon_0^{-2}}$ is concentrated around $2\varepsilon V$. In other words, the number of vertices in connected components of size at least ε_0^{-2} is close to $2\varepsilon V$ **whp**. However, on the event $\{Z_{\geq\varepsilon_0^{-2}} \geq 1\}$, we have that

$$|\mathcal{C}_{\max}| \leq Z_{\geq\varepsilon_0^{-2}}, \tag{2.6}$$

and so it follows that

$$|\mathcal{C}_{\max}| \leq 2\varepsilon V(1 + o_{\mathrm{p}}(1)). \tag{2.7}$$

■

The following result is an easy corollary to Proposition 2.2. It shows that, in fact, concentration of measure holds for the number of vertices in clusters of size at least ηV, for *all* $\eta \ll \varepsilon$ such that $\eta^3 V \gg 1$. This will be required for the proof of the lower bound of Theorem 1.1, as discussed in the proof overview above.

Corollary 2.3 (Concentration of the number of vertices in all large components)**.** *Set* $p = p_c + \frac{\varepsilon}{\Omega}$ *and let* $V^{-1/3}(\log V)^{1/3} \ll \varepsilon \ll 1$. *Let* $\eta = \eta(n)$ *satisfy* $\eta \ll \varepsilon$ *and* $\eta^3 V \gg 1$. *Then, for every* $\delta > 0$,

$$\mathbb{P}_p\Big(|Z_{\geq\eta V} - \mathbb{E}_p[Z_{\geq\eta V}]| \geq \delta\varepsilon V\Big) = o(1). \tag{2.8}$$

Corollary 2.3 allows us to use the concentration of $Z_{\geq\eta V}$ for any appropriate η, thus effectively removing the delicate choice of η in Proposition 2.2. We shall see that Corollary 2.3 follows from Propositions 2.1 and 2.2, combined with a simple first moment estimate.

Proof. Let us choose η satisfying both $\eta \ll \varepsilon$ and $\eta^3 V \gg 1$, so that in particular $\eta V \gg \eta^{-2} \gg \varepsilon^{-2}$. Choose further ε_0 as given in Proposition 2.2. We shall assume that $\varepsilon_0^{-2} \leq \eta V$; the proof when $\varepsilon_0^{-2} > \eta V$ is a simple adaptation of the argument below. By Proposition 2.1, for any fixed $\delta > 0$,

$$\begin{aligned}\mathbb{P}_p\Big(|Z_{\geq\eta V} - 2\varepsilon V| \geq \delta\varepsilon V\Big) &\leq \mathbb{P}_p\Big(|Z_{\geq\varepsilon_0^{-2}} - \mathbb{E}_p[Z_{\geq\varepsilon_0^{-2}}]| \geq \delta\varepsilon V/3\Big)\\ &\quad + \mathbb{P}_p\Big(|Z_{\geq\eta V} - Z_{\geq\varepsilon_0^{-2}}| \geq \delta\varepsilon V/3\Big),\end{aligned} \tag{2.9}$$

provided that V is large enough. More precisely, the volume V must be such that

$$\Big|\mathbb{E}_p[Z_{\geq\varepsilon_0^{-2}}] - 2\varepsilon V\Big| \leq \delta\varepsilon V/3, \tag{2.10}$$

or, equivalently (using transitivity),

$$\Big|\mathbb{P}_p(|C(\mathbf{v})| \geq \varepsilon_0^{-2}) - 2\varepsilon\Big| \leq \delta\varepsilon/3. \tag{2.11}$$

By Proposition 2.2,

$$\mathbb{P}_p\Big(|Z_{\geq\varepsilon_0^{-2}} - \mathbb{E}_p[Z_{\geq\varepsilon_0^{-2}}]| \geq \delta\varepsilon V/3\Big) = o(1). \tag{2.12}$$

Further, since $\varepsilon_0^{-2} \gg \varepsilon^{-2}$, the Markov inequality, together with the fact that $\varepsilon_0 \leq \eta$, yields that, for every $\delta > 0$,

$$\begin{aligned}\mathbb{P}_p\Big(|Z_{\geq\eta V} - Z_{\geq\varepsilon_0^{-2}}| \geq \delta\varepsilon V/3\Big) &\leq \frac{3\mathbb{E}_p[Z_{\geq\varepsilon_0^{-2}} - Z_{\geq\eta V}]}{\delta\varepsilon V} \\ &= \frac{3}{\delta\varepsilon}\Big[\mathbb{P}_p(|C(\mathbf{v})| \geq \varepsilon_0^{-2}) - \mathbb{P}_p(|C(\mathbf{v})| \geq \eta V)\Big] = o(1),\end{aligned} \tag{2.13}$$

where the last equality follows from Proposition 2.1 together with the fact that $\varepsilon^{-2} \ll \eta V \ll \varepsilon V$ and $\varepsilon^{-2} \ll \varepsilon_0^{-2} \ll \varepsilon V$. Equation (2.13) thus completes the proof. □

It remains to establish a matching lower bound for $|\mathcal{C}_{\max}|$, and we shall do this via a "sprinkling" argument. (Sprinkling is sometimes referred to as the "two-round exposure", see [18, Chapter 1].) This part of our proof is based on two results below. Before we state them, we need to introduce some more notation.

For each i, the *i-th horizontal line* of $H(2,n)$ is defined to be the set $\{(i,x) : x = 1,\dots,n\}$ of vertices with first coordinate i; similarly the set $\{(x,i) : x = 1,\dots,n\}$ of vertices with the second coordinate equal to i constitutes the *i-th vertical line*. A vertex belonging to a given line is said to be an *element* of that line.

Proposition 2.4 (Lower bound on the number of line elements in a large cluster)**.** *Set $p = p_c + \frac{\varepsilon}{\Omega}$. There exists a constant $C > 0$ such that the following holds. Fix ε, η satisfying $V^{-1/3} \ll \varepsilon \ll 1$, $\eta \ll \varepsilon$, $\eta V \gg \varepsilon^{-2}$ and $\eta V/n \geq C \log n$ for n sufficiently large. Then* ***whp*** *for every cluster of size at least ηV, there are at least $\frac{3n}{4}$ horizontal lines each with at least $\eta V/(4n)$ elements contained in the cluster.*

The proof of Proposition 2.4 is deferred to Section 4.3. Assuming it holds, we now prove that the second round exposure will join together every pair of large clusters formed during the first round. In the following lemma, for a pair of sets of vertices S_1, S_2, we use the notation $S_1 \longleftrightarrow S_2$ to denote the event that S_1, S_2 are joined together. We also write $S_1 \not\longleftrightarrow S_2$ to denote that S_1, S_2 are not joined together.

Lemma 2.5 (Sprinkling)**.** *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Choose $V^{-1/3} \ll \varepsilon \ll 1$. Let $\eta = \sqrt{\varepsilon}V^{-1/6}$, and let S_1, S_2 be disjoint sets of vertices both containing at least $\eta V/(4n)$ elements of at least $3n/4$ horizontal lines (possibly different lines for S_1 and S_2). Then*

$$\mathbb{P}_{\eta/\Omega}(S_1 \longleftrightarrow S_2) \geq 1 - o\Big(\frac{V^{-1/3}}{\varepsilon}\Big). \tag{2.14}$$

Proof. Choose two disjoint vertex sets S_1 and S_2 each containing at least $\eta V/4$ elements in at least $3n/4$ horizontal lines. Then S_1 and S_2 must have at least $n/2$ such lines in common, that is *both* S_1 and S_2 contain at least $\eta V/(4n) = \eta n/4$ elements of these lines. Note that, since $\varepsilon \gg V^{-1/3}$, $\eta = \sqrt{\varepsilon}V^{-1/6}$ and $V = n^2$, we have $\eta V/n = \sqrt{\varepsilon}V^{1/3} \gg 1$. Along the shared good lines, there are at least $(\eta n)^2/16$ edges with one endpoint in S_1 and the other in S_2. All of these edges will be occupied independently under $\mathbb{P}_{\eta/\Omega}$, so

$$\mathbb{P}_{\eta/\Omega}(S_1 \not\longleftrightarrow S_2) \leq \Big(1 - \frac{\eta}{2(n-1)}\Big)^{n(\eta n)^2/32}. \tag{2.15}$$

Using the inequality $1-x \leq e^{-x}$ and the fact that $\varepsilon \gg V^{-1/3}$,

$$\mathbb{P}_{\eta/\Omega}(S_1 \longleftrightarrow\!\!\!\!\!\!/\ \ S_2) \leq e^{-\eta^3 n^2/64} = e^{-\varepsilon^{3/2} V^{1/2}/64} \ll \Big(\varepsilon V^{1/3}\Big)^{-1} = o(1), \tag{2.16}$$

which completes the proof. □

We can now do the lower bound part of Theorem 1.1.

Proof of the lower bound part of Theorem 1.1. We choose $\eta = \sqrt{\varepsilon} V^{-1/6}$, as in Lemma 2.5, and note that the results in Proposition 2.4 apply to this choice of η. Indeed, since $\varepsilon \gg V^{-1/3}$, we have that $\eta = \varepsilon/(\sqrt{\varepsilon V^{1/3}}) \ll \varepsilon$ and $\eta V = \sqrt{\varepsilon} V^{5/6} \gg \varepsilon^{-2}$. Finally, once again using $\varepsilon \gg V^{-1/3}$, we have $\eta V/n = \sqrt{\varepsilon} V^{1/3} \gg V^{1/6} \geq C \log n$ for n sufficiently large.

We define p_- by the relation

$$p_- + (1-p_-)\frac{\eta}{\Omega} = p. \tag{2.17}$$

Note that every configuration with edge probability p can be obtained in a unique way as follows. First construct a configuration by throwing in edges independently of one another with probability p_-; subsequently, "sprinkle" extra edges with probability $\frac{\eta}{\Omega}$, independently of one another and of the p_- configuration. In the final configuration, an edge is occupied precisely when it is occupied in either the p_- configuration, or when it is an edge that is added during the sprinkling procedure. Since $\eta \ll \varepsilon$,

$$p_- = p + o\Big(\frac{\varepsilon}{n}\Big). \tag{2.18}$$

Let $Z'_{\geq \eta V}$ denote the number of vertices in connected components of size at least ηV in the p_- configuration. Since δ in Corollary 2.3 is arbitrary, it implies that $Z'_{\geq \eta V} = 2\varepsilon V(1+o_{\mathrm{p}}(1))$ after the first round of exposure; and by Proposition 2.4 **whp** every cluster of size at least ηV includes at least $\eta V/(4n)$ elements in at least $\frac{3n}{4}$ lines. Thus, under the measure $\mathbb{P}_{p_-}$, **whp** there are at most $\frac{2\varepsilon}{\eta}(1+o(1))$ connected clusters of size at least ηV, and each of these connected components contains at least $\eta V/(4n)$ elements in at least $\frac{3n}{4}$ lines.

It now suffices to prove that **whp** the subsequent sprinkling procedure (second round of exposure) joins together every pair of clusters of size at least ηV. Indeed, if this is the case, then after the sprinkling we end up with a single connected component of size at least

$$Z'_{\geq \eta V} \geq 2\varepsilon V(1+o_{\mathrm{p}}(1)). \tag{2.19}$$

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two vertices such that $C(\mathbf{v}_1) \neq C(\mathbf{v}_2)$, and $|C(\mathbf{v}_1)| \geq \eta V$ and $|C(\mathbf{v}_2)| \geq \eta V$. Let us take $S_1 = C(\mathbf{v}_1)$ and $S_2 = C(\mathbf{v}_2)$. By Proposition 2.4, we may assume that for both S_1 and S_2 one can find at least $\frac{3n}{4}$ lines (not necessarily the same ones for S_1 and S_2) each with at least $\eta V/(4n)$ elements in S_1 and S_2. Then, by Lemma 2.5,

$$\mathbb{P}_{\eta/\Omega}(S_1 \longleftrightarrow\!\!\!\!\!\!/\ \ S_2) = o\Big(\frac{V^{-1/3}}{\varepsilon}\Big). \tag{2.20}$$

But **whp** there are at most $\left(\frac{2\varepsilon}{\eta}\right)^2(1+o(1)) = O(\varepsilon V^{1/3})$ distinct choices for $C(\mathbf{v}_1)$ and $C(\mathbf{v}_2)$ with $|C(\mathbf{v}_1)| \geq \eta V$ and $|C(\mathbf{v}_2)| \geq \eta V$, and so a simple union bound implies that after sprinkling **whp** all connected components of size at least ηV are connected. By (2.19), this completes the proof. ■

The remainder of the paper is organised as follows. In Section 3 we prove auxiliary results relating to the tails of the total progeny of binomial Galton-Watson processes. Section 4 contains proofs of Propositions 2.1 and 2.4; therein we investigate the structure of percolation clusters (cluster tails and the number of elements per coordinate line) by comparing them to binomial

Galton-Watson processes. Finally, in Section 5, we establish concentration of measure for the number of vertices in large clusters, thus proving Proposition 2.2.

3. Total progeny of a Galton-Watson process

This section brings together some useful results from the theory of branching processes, which will play a key role at various stages in our proofs.

We consider a standard Galton-Watson process whose offspring distribution Z is a binomial $\mathrm{Bi}(N,p)$, where $N \in \mathbb{N}$ and $p \in [0,1]$ is the Hamming graph edge probability. We assume that with probability 1 the process begins with one individual. We write $\mathbb{P}_{N,p}$ for the probability measure corresponding to this process (implicitly assuming an underlying sample space and σ-field).

Let F be the total progeny or family size. Our aim is to prove the following three results concerning the distribution of F. Proposition 3.1 compares the distribution of F under the measures $\mathbb{P}_{N,p}$, $\mathbb{P}_{\tilde{N},p}$ for different values of N and $\tilde{N}$. Proposition 3.2 estimates the probability that the value of F is between ℓ and 2ℓ, for some (large) integer ℓ. Proposition 3.3 estimates the probability that F takes a value at least ℓ for some large integer ℓ.

Proposition 3.1 (Tails of total progeny in two binomial branching processes)**.** *Let $\ell \in \mathbb{N}$. Suppose that $N \in \mathbb{N}$ and $\tilde{N} = \tilde{N}(N)$ satisfy $N \geq \tilde{N}$. Further, assume that $\varepsilon = Np - 1$ is such that $\varepsilon \to 0$ and $\varepsilon \geq N^{-2/3}$; and that $\tilde{\varepsilon} = \tilde{N}p - 1 > 0$ and $|\varepsilon - \tilde{\varepsilon}| = o(\varepsilon)$ as $N \to \infty$. Then, for some constant $C > 0$, as $N \to \infty$,*

$$|\mathbb{P}_{N,p}(F \geq \ell) - \mathbb{P}_{\tilde{N},p}(F \geq \ell)| \leq C\Big(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{N\ell^{1/2}} + \frac{1}{\ell^3}\Big). \tag{3.1}$$

Proposition 3.2 (Bounds on the total progeny distribution)**.** *Let $N, \ell \in \mathbb{N}$. Suppose that $\varepsilon = Np - 1$ satisfies $\varepsilon \to 0$ and $\varepsilon \geq N^{-2/3}$ as $N \to \infty$. Then, for some constant $C > 0$, as $N \to \infty$,*

$$\mathbb{P}_{N,p}(F \in [\ell, 2\ell]) \leq \frac{C}{\sqrt{\ell}}. \tag{3.2}$$

Proposition 3.3 (Tails of the total progeny near criticality)**.** *Let $N, \ell \in \mathbb{N}$. Suppose that $\varepsilon = Np - 1$ satisfies $\varepsilon \to 0$ and $\varepsilon \geq N^{-2/3}$ as $N \to \infty$. Then, as $N \to \infty$,*

$$\mathbb{P}_{N,p}(F \geq \ell) = 2\varepsilon + O(\varepsilon^2) + O\Big(\frac{1}{\sqrt{\ell}}\Big). \tag{3.3}$$

We note that with a little care and minor modifications, Propositions 3.1–3.3 could be extended to the case where ε is a positive constant (i.e. strictly above the critical window). In the corresponding statement, (3.3) would have $\zeta_{1+\varepsilon}$ instead of 2ε, where, as before, ζ_λ denotes the survival probability of a Poisson Galton-Watson process with mean family size λ.

Our proofs of Propositions 3.1 and 3.2 will make use of the well-known *Otter-Dwass formula*, which describes the distribution of the total progeny of a branching process, see [11, 23]. We begin by stating a special case of this formula (due to Otter) for a branching process starting with 1 individual. (The formula was later extended by Dwass to a process starting with r individuals, for arbitrary $r \in \mathbb{N}$, but we do not make use of the extension here.)

Lemma 3.4 (Otter-Dwass formula)**.** *Let $Z_1, Z_2, Z_3, \ldots$ be i.i.d. random variables distributed as Z. Let $\mathbb{P}$ denote the Galton-Watson process measure. For all $k \in \mathbb{N}$,*

$$\mathbb{P}(F = k) = \frac{1}{k}\mathbb{P}(\sum_{i=1}^{k} Z_i = k - 1).$$

We now prove each of Propositions 3.1–3.3 in turn.

Proof of Proposition 3.1. It will be convenient for us to introduce new parameters $\lambda = pN$ and $\tilde{\lambda} = p\tilde{N}$. By assumption $\lambda, \tilde{\lambda} > 1$, so that both branching processes are supercritical. The total progeny size in the $\mathrm{Bi}(\tilde{N}, p)$ process will be denoted by $\tilde{F}$.

By Lemma 3.4, for each $k \in \mathbb{N}$,

$$\mathbb{P}_{N,p}(F = k) = \frac{1}{k}\mathbb{P}(\sum_{i=1}^{k} Z_i = k-1), \tag{3.4}$$

where the Z_i are i.i.d. $\mathrm{Bi}(N, \lambda/N)$. We now investigate the asymptotics of the formula (3.4) as $N \to \infty$, for large integers $k = k(N)$. Our aim is to obtain estimates for $\mathbb{P}_{N,p}(F = k)$ sharp enough for the errors to be summable.

In outline, our calculation is as follows. First we demonstrate that, if $k \geq C_0\varepsilon^{-2}\log(1/\varepsilon)$ for a sufficiently large constant C_0, then

$$\mathbb{P}_{N,p}(k \leq F < \infty) \leq k^{-4}. \tag{3.5}$$

Next we show that (3.5) implies an identical upper bound for $\mathbb{P}_{\tilde{N},p}(\tilde{F} = k)$. Subsequently, we prove that there exists a constant $\tilde{C} > 0$ such that, if $k \leq C_0\varepsilon^{-2}\log(1/\varepsilon)$, then

$$|\mathbb{P}_{N,p}(F = k) - \mathbb{P}_{\tilde{N},p}(\tilde{F} = k)| \leq \frac{\tilde{C}}{k^{3/2}}\exp\Big(-\frac{(k-1)\varepsilon^2}{4}\Big)\Big|(1+k\varepsilon)|\varepsilon - \tilde{\varepsilon}| + \frac{k}{N^3} + \frac{\varepsilon}{N} + \frac{1}{kN}\Big|, \tag{3.6}$$

We can then sum the errors in (3.6) and (3.5) to show that, for any $\ell \in \mathbb{N}$,

$$|\mathbb{P}_{N,p}(\ell \leq F < \infty) - \mathbb{P}_{\tilde{N},p}(\ell \leq \tilde{F} < \infty)| \leq C\Big(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{N\ell^{1/2}} + \frac{1}{\ell^3}\Big). \tag{3.7}$$

Finally, since $\lambda > 1$, we need to estimate $\mathbb{P}_{N,p}(F = \infty)$ and $\mathbb{P}_{\tilde{N},p}(\tilde{F} = \infty)$; we shall show that

$$\big|\mathbb{P}_{N,p}(F = \infty) - \mathbb{P}_{\tilde{N},p}(\tilde{F} = \infty)\big| \leq C|\varepsilon - \tilde{\varepsilon}|. \tag{3.8}$$

Combining the last two estimates yields Proposition 3.1. As we shall see below, several steps of our proof will also play a role in proving Propositions 3.2 and 3.3.

Let us make a start on the details. To show (3.5), we note that (3.4) implies

$$\mathbb{P}_{N,p}(F = k) \leq \frac{1}{k}\mathbb{P}(\sum_{i=1}^{k} Z_i \leq k), \tag{3.9}$$

where $\sum_{i=1}^{k} Z_i \sim \mathrm{Bi}(kN, p)$. But if $Z \sim \mathrm{Bi}(n, p)$, then (see for instance [16])

$$\mathbb{P}(Z \leq np - t) \leq e^{-\frac{t^2}{2(np + \frac{t}{3})}}. \tag{3.10}$$

Applying (3.10) with $n = kN$, $p = \frac{1+\varepsilon}{N}$ and $t = np - k = k\varepsilon$ (and also using our assumption that $\varepsilon < 1$), we obtain that

$$\mathbb{P}_{N,p}(F = k) \leq \frac{1}{k}e^{-\frac{k\varepsilon^2}{2(1+\frac{4\varepsilon}{3})}} \leq \frac{1}{k}e^{-\frac{k\varepsilon^2}{5}}. \tag{3.11}$$

But if $k \geq C_0\varepsilon^{-2}\log(1/\varepsilon)$ for a sufficiently large $C_0 > 0$ then

$$\frac{1}{k}e^{-\frac{k\varepsilon^2}{4}} \leq k^{-4}, \tag{3.12}$$

so (3.5) follows.

We now show that (3.6) holds for $k \leq C_0\varepsilon^{-2}\log(1/\varepsilon)$. Clearly, (3.4) implies that, for each $k \in \mathbb{N}$,

$$\mathbb{P}_{N,p}(F=k) = \frac{1}{k}\binom{kN}{k-1}\Big(\frac{\lambda}{N}\Big)^{k-1}\Big(1-\frac{\lambda}{N}\Big)^{kN-k+1}. \tag{3.13}$$

By Stirling's formula,

$$(m)_r := m(m-1)\cdots(m-r+1) = m^r \exp\Big(-\frac{r^2}{2m}-\frac{r^3}{6m^2}+O\Big(\frac{r^4}{m^3}\Big)\Big).$$

Applying the above approximation with $m = kN$ and $r = k-1$ (and noting that $\frac{r^4}{m^3} = O(\frac{k}{N^3})$), we arrive at

$$\begin{aligned}\mathbb{P}_{N,p}(F=k) &= \frac{1}{k}\frac{(k\lambda)^{k-1}}{(k-1)!}\exp\Big(-\frac{(k-1)^2}{2kN}-\frac{(k-1)^3}{6k^2N^2}+O\Big(\frac{k}{N^3}\Big)\Big)\Big(1-\frac{\lambda}{N}\Big)^{kN-k+1}\\ &= \frac{(k\lambda)^{k-1}}{k!}\exp\Big(-\frac{k}{2N}+\frac{1}{N}-\frac{1}{2kN}-\frac{k}{6N^2}+O\Big(\frac{1}{N^2}+\frac{k}{N^3}\Big)\Big)\Big(1-\frac{\lambda}{N}\Big)^{kN-k+1}.\end{aligned}$$

Observe that

$$\begin{aligned}\Big(1-\frac{\lambda}{N}\Big)^{kN-k+1} &= \exp\big((kN-k+1)\log(1-\frac{\lambda}{N})\big)\\ &= \exp\Big(-\lambda k+\frac{k\lambda}{N}-\frac{\lambda}{N}-\frac{\lambda^2 k}{2N}+\frac{\lambda^2 k}{2N^2}-\frac{\lambda^3 k}{3N^2}+O(\frac{1}{N^2}+\frac{k}{N^3})\Big),\end{aligned}$$

so that

$$\begin{aligned}\mathbb{P}_{N,p}(F=k) = \frac{k^{k-1}e^{-(k-1)}e^{-\lambda}}{k!}&\exp((k-1)f(\lambda))\exp\Big(-\frac{k(\lambda-1)^2}{2N}+\frac{1-\lambda}{N}-\frac{1}{2kN}\Big)\\ &\times\exp\Big(\frac{k}{N^2}g(\lambda)+O\Big(\frac{k}{N^3}+\frac{1}{N^2}\Big)\Big),\end{aligned} \tag{3.14}$$

where

$$f(\lambda) = \log\lambda-(\lambda-1), \qquad g(\lambda) = -\frac{1}{6}+\frac{\lambda^2}{2}-\frac{\lambda^3}{3}. \tag{3.15}$$

The Taylor expansion for $|\lambda-1|$ small gives

$$f(\lambda) = -\frac{(\lambda-1)^2}{2}+O(|\lambda-1|^3). \tag{3.16}$$

But $\lambda-1=\varepsilon$ and $k \leq C_0\varepsilon^{-2}\log(1/\varepsilon)$, where $N^{-2/3} \ll \varepsilon = o(1)$, and so $k|\lambda-1|^3 = o(1)$, uniformly for all such k. The other error terms can be bounded similarly, and hence, uniformly for $k \leq C_0\varepsilon^{-2}\log(1/\varepsilon)$,

$$\mathbb{P}_{N,p}(F=k) = (1+o(1))\frac{k^{k-1}e^{-k}}{k!}\exp\Big(-\frac{1}{2}(k-1)(\lambda-1)^2\Big). \tag{3.17}$$

We now compare $\mathbb{P}_{N,p}(F=k)$ to $\mathbb{P}_{\tilde{N},p}(\tilde{F}=k)$ for $k \leq C_0\varepsilon^{-2}\log(1/\varepsilon)$. Write $\lambda/N = \tilde{\lambda}/\tilde{N}$, where $\tilde{\lambda} = \tilde{N}\lambda/N < \lambda$. Then a calculation similar to the one above shows that

$$\begin{aligned}\mathbb{P}_{\tilde{N},p}(\tilde{F}=k) = \frac{k^{k-1}e^{-\tilde{\lambda}-(k-1)}}{k!}&\exp\Big((k-1)f(\tilde{\lambda})\Big)\\ &\times\exp\Big(-\frac{k(\tilde{\lambda}-1)^2}{2\tilde{N}}+\frac{1-\tilde{\lambda}}{\tilde{N}}-\frac{1}{2k\tilde{N}}+\frac{k}{\tilde{N}^2}g(\tilde{\lambda})+O\Big(\frac{k}{\tilde{N}^3}+\frac{1}{\tilde{N}^2}\Big)\Big)\Big).\end{aligned}$$

By assumption, $\varepsilon = \lambda - 1 > 0$ and $\tilde{\varepsilon} = \tilde{\lambda} - 1 > 0$. Further,

$$f(\lambda) - f(\tilde{\lambda}) = (\tilde{\varepsilon} - \varepsilon) f'(1+s), \tag{3.18}$$

for some $s \in (\tilde{\varepsilon}, \varepsilon)$, where

$$|f'(1+s)| = \frac{s}{1+s} \leq s, \quad s \geq 0, \tag{3.19}$$

and so

$$f(\lambda) - f(\tilde{\lambda}) = O(\varepsilon|\varepsilon - \tilde{\varepsilon}|). \tag{3.20}$$

We deduce that, for $\varepsilon \geq N^{-2/3}$ and all $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon)$,

$$|\mathbb{P}_{N,p}(F = k) - \mathbb{P}_{\tilde{N},p}(\tilde{F} = k)| = \frac{(k/e)^{k-1} e^{-\lambda}}{k!} \exp((k-1) f(\lambda)) \big| \exp(x) - \exp(y) \big| \tag{3.21}$$

where

$$\begin{aligned} x &= -\frac{k\varepsilon^2}{2N} + \frac{k}{N^2} g(\lambda) + O\Big(\frac{k}{N^3} + \frac{\varepsilon}{N} + \frac{1}{kN}\Big), \\ y &= (\varepsilon - \tilde{\varepsilon}) - \frac{k\tilde{\varepsilon}^2}{2\tilde{N}} + \frac{k}{\tilde{N}^2} g(\tilde{\lambda}) + O(k\varepsilon|\varepsilon - \tilde{\varepsilon}|) + O\Big(\frac{k}{\tilde{N}^3} + \frac{\tilde{\varepsilon}}{\tilde{N}} + \frac{1}{k\tilde{N}}\Big). \end{aligned}$$

Since $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon)$, $N^{-2/3} \ll \varepsilon = o(1)$ and $\tilde{\varepsilon} = o(\varepsilon)$, $x = o(1)$ and also all contributions to y are $o(1)$, except possibly for the term $k\varepsilon|\varepsilon - \tilde{\varepsilon}|$. Now, for some constant C,

$$|\exp(x) - \exp(y)| \leq C|x - y| e^{|x| \vee |y|}, \tag{3.22}$$

where, for $u, v \in \mathbb{R}$, $u \vee v = \max\{u, v\}$. Note that $x = o(1)$ and, since $|\varepsilon - \tilde{\varepsilon}| = o(\varepsilon)$, we have that $y = o(1) + o(k\varepsilon^2)$. As a result, we obtain that, for N sufficiently large,

$$|\exp(x) - \exp(y)| \leq C|x - y| e^{(k-1)\varepsilon^2/4}. \tag{3.23}$$

Since further $N^{-2} - \tilde{N}^{-2} = O(N^{-2}|\varepsilon - \tilde{\varepsilon}|)$, the contribution to $|x-y|$ due to the term $g(\lambda)k/N^2 - g(\tilde{\lambda})k/\tilde{N}^2$ is $O(kN^{-2}|\varepsilon - \tilde{\varepsilon}|) = o(k\varepsilon|\varepsilon - \tilde{\varepsilon}|)$, which gives

$$|x - y| \leq C(1 + k\varepsilon)|\varepsilon - \tilde{\varepsilon}| + O\Big(\frac{k}{N^3} + \frac{\varepsilon}{N} + \frac{1}{kN}\Big). \tag{3.24}$$

Hence, combining (3.21), (3.23), (3.16), for all $k \leq C_0 \varepsilon^{-2} \log(1/\varepsilon^3)$, we arrive at (3.6).

Summing the estimates (3.5) and (3.6) over $k \geq \ell$,

$$\begin{aligned} &|\mathbb{P}_{N,p}(\ell \leq F < \infty) - \mathbb{P}_{\tilde{N},p}(\ell \leq \tilde{F} < \infty)| \\ &\quad \leq C \sum_{k \geq \ell} \Big((1 + k\varepsilon)|\varepsilon - \tilde{\varepsilon}| + \frac{k}{N^3} + \frac{\varepsilon}{N} + \frac{1}{kN}\Big) \frac{\exp(-(k-1)\varepsilon^2/4)}{k^{3/2}} + \sum_{k \geq \ell} k^{-4}. \end{aligned}$$

The final contribution is $O(\ell^{-3})$; for the remaining terms observe that

$$\sum_{k \geq \ell} k^{-a} \exp(-(k-1)\varepsilon^2/4) \leq \begin{cases} C\ell^{1-a} & \text{for } a > 1, \\ C\varepsilon^{-2-2a} & \text{for } a < 1, \end{cases} \tag{3.25}$$

which yields (with a suitably adjusted value of C)

$$\begin{aligned} |\mathbb{P}_{N,p}(\ell \leq F < \infty) - \mathbb{P}_{\tilde{N},p}(\ell \leq \tilde{F} < \infty)| &\leq C\Big(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{\varepsilon N^3} + \frac{\varepsilon}{N\ell^{1/2}} + \frac{1}{\ell^{3/2} N} + \frac{1}{\ell^3}\Big) \\ &\leq C\Big(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{N\ell^{1/2}} + \frac{1}{\ell^3}\Big). \end{aligned} \tag{3.26}$$

To prove (3.8), we need to estimate $|a - \tilde{a}|$, where $a = \mathbb{P}_{N,p}(F < \infty)$ and $\tilde{a} = \mathbb{P}_{\tilde{N},p}(\tilde{F} < \infty)$. The quantities $a, \tilde{a}$ respectively are the smallest positive roots of the equations

$$a = (1 + \frac{\lambda}{N}(a-1))^N, \qquad \text{and} \qquad \tilde{a} = (1 + \frac{\lambda}{N}(\tilde{a}-1))^{\tilde{N}}. \tag{3.27}$$

Using the convexity of probability generating functions and the supercriticality of the branching processes in question, the equations in (3.27) each have precisely one root a and $\tilde{a}$ respectively in the interval $[0, 1)$.

The proof is divided into two main steps. In the first step, we prove that $1 - a = 2\varepsilon + O(\varepsilon^2)$, which also implies that $|a - \tilde{a}| = o(\varepsilon)$ when $|\varepsilon - \tilde{\varepsilon}| = o(\varepsilon)$, so that we may use a Taylor expansion. In the second main step, we prove that $|a - \tilde{a}| \le C|\varepsilon - \tilde{\varepsilon}|$.

To prove that $1 - a = 2\varepsilon + O(\varepsilon^2)$, we expand the right hand side of (3.27) to obtain

$$\begin{aligned} a - 1 &= \binom{N}{1}\frac{\lambda}{N}(a-1) + \binom{N}{2}\Big(\frac{\lambda}{N}\Big)^2(a-1)^2 + O(|a-1|^3) \\ 1 &= \lambda + \frac{N-1}{2N}\lambda^2(a-1) + O(|a-1|^2) \\ 1 &= 1 + \varepsilon + \frac{N-1}{2N}(1+\varepsilon)^2(a-1) + O(|a-1|^2), \end{aligned}$$

so that

$$(1-a)(1 + 2\varepsilon + \varepsilon^2 - N^{-1} - 2N^{-1}\varepsilon + O(N^{-1}\varepsilon^2)) = 2\varepsilon + O(|1-a|^2),$$

and so, again using that $\varepsilon \ge N^{-2/3}$,

$$1 - a = 2\varepsilon + O(\varepsilon^2). \tag{3.28}$$

To prove that $|a - \tilde{a}| \le C|\varepsilon - \tilde{\varepsilon}|$, we use that

$$a - \tilde{a} = (1+\frac{\lambda}{N}(a-1))^N - (1+\frac{\lambda}{N}(\tilde{a}-1))^{\tilde{N}} = f_{\tilde{N}}(a) - f_{\tilde{N}}(\tilde{a}) + f_{\tilde{N}}(a)\Big((1+\frac{\lambda}{N}(a-1))^{N-\tilde{N}} - 1\Big), \tag{3.29}$$

where $f_{\tilde{N}}(x) = (1 + \frac{\lambda}{N}(x-1))^{\tilde{N}}$. Note first that

$$(1 + \frac{\lambda}{N}(a-1))^{N-\tilde{N}} - 1 = O\Big(\frac{N-\tilde{N}}{N}(1-a)\Big). \tag{3.30}$$

Further, since $|a - \tilde{a}| = o(|a-1|)$, we have that $f_{\tilde{N}}(a) \le 2$. Also,

$$f'_{\tilde{N}}(x) = \frac{\tilde{N}\lambda}{N}\Big(1 + \frac{\lambda}{N}(x-1)\Big)^{\tilde{N}-1}, \tag{3.31}$$

$$f''_{\tilde{N}}(x) = \frac{\tilde{N}\lambda}{N}\frac{(\tilde{N}-1)\lambda}{N}\Big(1 + \frac{\lambda}{N}(x-1)\Big)^{\tilde{N}-2}, \tag{3.32}$$

and it is not hard to see that $f''_{\tilde{N}}(x) = O(1)$ uniformly for $x \in [\tilde{a}, a]$. Hence

$$a - \tilde{a} = (a - \tilde{a})f'_{\tilde{N}}(\tilde{a}) + O\Big(\frac{N-\tilde{N}}{N}(1-a)\Big) + O\big((a-\tilde{a})^2\big), \tag{3.33}$$

so that

$$a - \tilde{a} = O\Big(\frac{(N-\tilde{N})(1-a)}{N|1 - f'_{\tilde{N}}(\tilde{a})|}\Big) + O\Big(\frac{(a-\tilde{a})^2}{|1 - f'_{\tilde{N}}(\tilde{a})|}\Big). \tag{3.34}$$

A closer inspection of $f'_{\tilde{N}}(x)$ yields that $f'_{\tilde{N}}(\tilde{a})-1=\varepsilon+o(\varepsilon)$, so that

$$|a-\tilde{a}|\leq O\Big(\frac{|N-\tilde{N}|}{N}\Big)+O\Big(\frac{(a-\tilde{a})^2}{1-a}\Big)\leq C|\varepsilon-\tilde{\varepsilon}|. \tag{3.35}$$

This completes the proof of Proposition 3.1. ∎

Proof of Proposition 3.2. By (3.17), for all $k\leq C_0\varepsilon^{-2}\log(1/\varepsilon)$

$$\mathbb{P}_{N,p}(F=k)=(1+o(1))\frac{(k/e)^{k-1}e^{-\lambda}}{k!}\exp\Big(-\frac{1}{2}(k-1)\varepsilon^2\Big).$$

Also (provided C_0 is large enough) for $k\geq C_0\varepsilon^{-2}\log(1/\varepsilon)$,

$$\mathbb{P}_{N,p}(F=k)\leq k^{-4}.$$

Summing over $\ell\leq k\leq 2\ell$ we obtain

$$\mathbb{P}_{N,p}(\ell\leq F\leq 2\ell)\leq C\sum_{\ell\leq k\leq 2\ell}\Big(\frac{1}{k^{3/2}}e^{-k\varepsilon^2/2}+k^{-4}\Big)\leq\frac{C}{\ell^{1/2}},$$

where the constant C was adjusted within the final inequality. ∎

Proof of Proposition 3.3. We have

$$\mathbb{P}_{N,p}(F\geq\ell)=1-\mathbb{P}_{N,p}(F<\infty)+\mathbb{P}_{N,p}(\ell\leq F<\infty). \tag{3.36}$$

By (3.28), the term $1-\mathbb{P}_{N,p}(F<\infty)$ is $2\varepsilon+O(\varepsilon^2)$. Calculations similar to those in the proof of Proposition 3.2 show that the final term is bounded by $O(\ell^{-1/2})$, which completes the proof. ∎

4. Comparisons to branching processes

In this section, we use comparisons to branching processes and concentration of measure techniques to study the cluster tail probabilities (cf. Proposition 2.1), as well as the cluster structure, specifically, the number of vertices per line, of large clusters (cf. Proposition 2.4).

This section is organised as follows. In Section 4.1, we describe a *cluster exploration* procedure, state key estimates for the tails of the cluster size distribution, and prove the upper bound part of Proposition 2.1. In Section 4.2 we establish an upper bound on the number of elements per line in a large cluster; this result is a crucial ingredient in the proof of Proposition 2.4. Section 4.3 contains a proof of the lower bound part of Proposition 2.1, as well as a proof of Proposition 2.4.

4.1. **Component exploration and strategy of proof.** We take an initial vertex $\mathbf{v}_0=(x_0,y_0)$ and *explore* its cluster, $C(\mathbf{v}_0)$, by *exploring* the vertices in that cluster successively one at a time, in a breadth-first order. Exploring a vertex (x,y) means that we consider all the edges (x,j) for $j\neq y$ in the order of increasing j, and decide for each one in turn if it is open with probability p or closed with probability $1-p$; then we do the same for the edges (i,y) for $i\neq x$ in the order of increasing i. Note that, until the moment all available vertices in the cluster have been explored, the number of explored vertices at time t is equal to t.

Let us introduce colours as follows. At time t, all vertices that have not yet been explored and are not yet contained in $C(\mathbf{v}_0)$ are *white*. All unexplored vertices connected to $\mathbf{v}_0$ (that is, included in $C(\mathbf{v}_0)$) at time t are *green*. All explored vertices are *red*. (Thus, in particular, at time 0 all vertices are white except for $\mathbf{v}_0$, which is green.) In fact, we need to modify this exploration process slightly as follows: when exploring a green vertex we only consider those of its edges where the other endpoint of the edge is white. If such an edge is found to be open, then we colour its other endpoint green.

Let $C_t(\mathbf{v}_0)$ be the set of vertices included in the cluster of $\mathbf{v}_0$ by the time t. Let also $G_t(\mathbf{v}_0)$ be the set of green vertices in the cluster at time t. Thus $C_t(\mathbf{v}_0)$ consists of all green and red vertices at time t, and $R_t(\mathbf{v}_0) = C_t(\mathbf{v}_0) \setminus G_t(\mathbf{v}_0)$ is the set of red vertices at time t. All the remaining vertices in the graph are white.

Let $T_{\mathbf{v}_0}$ denote the smallest time t when there are no green vertices remaining, that is $T_{\mathbf{v}_0} = \inf\{t : |G_t(\mathbf{v}_0)| = 0\}$. Note that $T_{\mathbf{v}_0} = |C(\mathbf{v}_0)|$, the size of the cluster of vertex $\mathbf{v}_0$, and $|R_t(\mathbf{v}_0)| = t$ for all $t \le T_{\mathbf{v}_0}$. Choose a parameter $\eta = \eta(\varepsilon, V)$ such that $0 < \eta \ll \varepsilon$ and let

$$T = T_{\mathbf{v}_0} \wedge \lceil \eta V \rceil \tag{4.1}$$

be the minimum of $T_{\mathbf{v}_0}$ and $\lceil \eta V \rceil$.

Given an integer $i \in \{1, \dots, n\}$, let $C_t(\mathbf{v}_0, i)$ be the set of vertices (i, y) included in the cluster at time t. (This is the collection of all the elements of the i-th horizontal line added by time t during the exploration procedure.) Let also $C(\mathbf{v}_0, i)$ be the set of all vertices (i, y) in $C(\mathbf{v}_0)$, that is, the collection of all the elements in the i-th horizontal line contained in $C(\mathbf{v}_0)$. We further denote the number of elements of the i-th horizontal line included in $C(\mathbf{v}_0)$ until time T by $N(\mathbf{v}_0, i) = |C_T(\mathbf{v}_0, i)|$.

Similarly, let $\hat{C}_t(\mathbf{v}_0, i)$ be the set of vertices (x, i) included in the cluster at time t, that is all the i-th vertical line elements added by time t during the exploration procedure.) Let $\hat{C}(\mathbf{v}_0, i)$ be the set of all vertices (x, i) in $C(\mathbf{v}_0)$; and, finally, denote the number of elements of the i-th vertical line included in $C(\mathbf{v}_0)$ until time T by $\hat{N}(\mathbf{v}_0, i) = |\hat{C}_T(\mathbf{v}_0, i)|$.

We write (x_t, y_t) for the vertex that is explored at time t if such a vertex exists, that is, if $t \le T$. We may identify the set of colours with the set $\{0, 1, 2\}$. The state of the exploration process at time t is the list giving the colour of each vertex, in other words, an n-vector with values in $\{0, 1, 2\}^n$. This process defines a natural filtration $\varphi_0 \subseteq \varphi_1 \subseteq \ldots \subseteq \varphi_T$, where φ_t is the smallest σ-field with respect to which the state at time t is measurable. (Informally, φ_t corresponds to "everything that has occurred until time t".) We note that T is a stopping time with respect to this filtration. We note also that, even on the event $\{T = \lceil \eta V \rceil\}$, it is not necessarily the case that $C_T(\mathbf{v}_0) = \lceil \eta V \rceil$, since the number of new vertices added at each exploration step is a random variable, which can be smaller or greater than 1. We stop our process at time T, and we make the convention that $C_t(\mathbf{v}_0) = C_T(\mathbf{v}_0)$ for all $t \ge T$ (and similarly for all other relevant random variables). This is important when $T = T_{\mathbf{v}_0} < \eta V$, that is, when the process dies out before time ηV.

Following the notation of Section 3, we let F denote the total population size of a Galton-Watson process starting with one individual, where the offspring distribution is $\mathrm{Bi}(\Omega, p)$; and further $\mathbb{P}_{\Omega,p}$ denotes the probability measure corresponding to this branching process. Proposition 2.1 involves upper and lower bounds on $\mathbb{P}_p(|C(\mathbf{v}_0)| \ge \ell)$ for appropriate choices of ℓ. These bounds are formulated in Lemmas 4.1– 4.3 below.

Lemma 4.1 (Stochastic domination of cluster size by branching process progeny size)**.** *For every* $\ell \in \mathbb{N}$,

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \ge \ell) \le \mathbb{P}_{\Omega,p}(F \ge \ell). \tag{4.2}$$

The result in Lemma 4.1 is standard, and we will omit its proof. In essence, it follows since in the cluster exploration, from each vertex being explored, *at most* $\mathrm{Bi}(\Omega, p)$ new vertices can be added to the cluster, independently of what has already been added. Thus the total cluster size must be at most the total population size of the binomial Galton-Watson process, as claimed.

Lemma 4.2 below follows directly from Lemma 4.1 and Proposition 3.3, and establishes the upper bound part of Proposition 2.1. It is also used in the proof of Lemma 5.2 in Section 5.

Lemma 4.2. *For every $\ell \in \mathbb{N}$, and for $\varepsilon \geq V^{-1/3}$,*

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \leq 2\varepsilon + O(\varepsilon^2) + O\Big(\frac{1}{\sqrt{\ell}}\Big).$$

In particular, if $\varepsilon^{-2} \ll \eta V \ll \varepsilon V$, then

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \eta V) \leq 2\varepsilon(1 + o(1)). \tag{4.3}$$

Proof. By Lemma 4.1, for every $\ell \in \mathbb{N}$,

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \leq \mathbb{P}_{\Omega,p}(F \geq \ell).$$

Our choice of $p = p(n)$ implies that $\Omega p = 1 + \varepsilon > 1$, that is, the $\mathrm{Bi}(\Omega, p)$ Galton-Watson process is supercritical. By Proposition 3.3,

$$\mathbb{P}_{\Omega,p}(F \geq \ell) = 2\varepsilon + O(\varepsilon^2) + O\Big(\frac{1}{\sqrt{\ell}}\Big).$$

For $\ell = \eta V$, we have that $\eta V \gg \varepsilon^{-2}$, and so $1/\sqrt{\eta V} = o(\varepsilon)$, which completes the proof. □

Our next lemma establishes a lower bound on $\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell)$, that is, the lower bound part of Proposition 2.1.

Lemma 4.3 (Stochastic domination of cluster size over branching process progeny size)**.** *For every $\ell \ll \varepsilon V$,*

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \geq \mathbb{P}_{\tilde{\Omega},p}(F \geq \ell) + O(V^{-3}), \tag{4.4}$$

where $\tilde{\Omega} = \Omega - \frac{5}{2}\max\{\ell n^{-1}, C \log n\}$.
Consequently, if $\varepsilon \gg V^{-1/3}$, $\eta \ll \varepsilon$ and $\varepsilon^{-2} \ll \eta V$, then

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \eta V) \geq 2\varepsilon(1 + o(1)). \tag{4.5}$$

Lemma 4.3 is proved in Section 4.3, where we show that the cluster size stochastically dominates the $\mathrm{Bi}(\tilde{\Omega}, p)$ Galton-Watson process.

4.2. **Upper bounds on the cluster size and structure.** In this section we give an upper bound on the number of elements of a large cluster that belong to a particular horizontal line. The following proposition is crucial in the proofs of Proposition 2.4 and Lemma 4.3.

Proposition 4.4 (Upper bound on the number of elements per line in a large cluster)**.** *Let $\varepsilon = \varepsilon(n) \geq 0$ be such that $\varepsilon = \varepsilon(n) \leq 1/20$ and choose $\eta \ll \varepsilon$. Further, let $N(\mathbf{v_0}, i)$ be the number of elements of $C_T(\mathbf{v_0}) = C_{\lceil \eta V \rceil}(\mathbf{v_0})$ that belong to the horizontal line i. There exists a positive constant c_1 such that for every $\nu > 0$*

$$\mathbb{P}_p\Big(\max_{i=1,\dots,n} : N(\mathbf{v_0}, i) \geq (1 + \nu)\frac{11}{9}\eta n\Big) \leq n e^{-c_1 \nu \eta n}. \tag{4.6}$$

Furthermore, there exist constants $c_2, c_3, c_4 > 0$ such that the following holds:

(1) *Let $n \in \mathbb{N}$ and $\eta = \eta(n)$ be such that $\eta n \geq c_2 \log n$. If n is sufficiently large, then*

$$\mathbb{P}_p\Big(\max_{i=1,\dots,n} : N(\mathbf{v_0}, i) \geq \frac{5}{4}\eta n\Big) \leq c_4 V^{-3}. \tag{4.7}$$

(2) *Let $n \in \mathbb{N}$ and $\eta = \eta(n)$ be such that $\eta n / \log n < c_2$. If n is sufficiently large, then*

$$\mathbb{P}_p\Big(\max_{i=1,\dots,n} : N(\mathbf{v_0}, i) \geq c_3 \log n\Big) \leq c_4 V^{-3}. \tag{4.8}$$

Here is an informal outline of the proof. Whenever we explore a vertex *not* on the line i, we add an element of line i with probability p. On the other hand, each vertex belonging to the line i has $n-1$ neighbours on that line. Whenever such a vertex is explored, each one of its neighbours on the line i is included with probability p (unless it is already in the cluster). It follows that the number of new elements on line i resulting from exploring a vertex belonging to that line is stochastically dominated by a $\mathrm{Bi}(n-1,p)$ Galton-Watson process. Since $p = \frac{1+\varepsilon}{2(n-1)}$ and $\varepsilon \le 1/2 < 1$ for n large enough, we have that $(n-1)p < 1$, so that the Galton-Watson process is subcritical. Hence, using standard concentration of measure techniques, we are able to upper bound the number of elements on line i that make it into a large cluster. We now make this argument precise.

Proof of Proposition 4.4. Let $i \in \{1,\dots,n\}$ and, for each $t = 1,\dots,T$, let $S_t(i)$ be the number of times s such that (x_{s-1}, y_{s-1}), (i, y_{s-1}) is open and $x_{s-1} \neq i$. That is, for each time $t \le T$, $S_t(i)$ is the number of times we enter the horizontal line i until time t. We can write

$$S_t(i) = \sum_{s=1}^{t} Y_s(i),$$

where $Y_t(i)$ is the indicator of the event that the edge between (x_{t-1}, y_{t-1}), (i, y_{t-1}) is open, and $x_{t-1} \neq i$. For each time t,

$$\mathbb{P}_p(Y_t(i) = 1 | \varphi_{t-1}) \le p,$$

and so known results (see for instance Lemma 2.2 in [20]) imply that $S_t(i)$ is stochastically dominated by a $\mathrm{Bi}(t,p)$ random variable. Consequently, for every $u \ge 0$, the moment generating function $M_{S_t(i)}(u)$ is bounded above by $(1 + p(e^u - 1))^t$.

For $r = 1, 2, \dots$, let $Z_r(i)$ be the number of vertices (i, x) added as a result of the r-th entry on to horizontal line i. Given that vertex $(i, \tilde{x}_0) \in C(\mathbf{v}_0)$, the number of its neighbours (i, x) added to $C(\mathbf{v}_0)$ during its exploration (if it has occurred by the time ηV) is easily seen to be stochastically dominated by a random variable $\mathrm{Bi}(n-1,p)$. Hence, for each r, $Z_r(i)$ is stochastically dominated by the total progeny in a branching process with offspring distribution $\mathrm{Bi}(n-1,p)$ descending from a single individual. Since $p = (1+\varepsilon)/2(n-1)$ and $\varepsilon < 1/2$, this branching process is subcritical. We deduce that, for $u \ge 0$, the moment generating function $M_{Z_r(i)}(u)$ of $Z_r(i)$ is bounded above by the moment generating function $M_Z(u)$ of an integer-valued, finite random variable Z, whose distribution is given by the Otter-Dwass formula (Lemma 3.4). In other words, for each $N \in \mathbb{N}$,

$$\mathbb{P}_p(Z = N) = \frac{\mathbb{P}(\xi_1 + \ldots + \xi_N = N - 1)}{N}, \tag{4.9}$$

where the ξ_r are i.i.d. $\mathrm{Bi}(n-1,p)$. It follows that

$$\begin{aligned} M_Z(u) &= \sum_{N=1}^{\infty} \frac{e^{uN}}{N} \mathbb{P}\big(\mathrm{Bi}(N(n-1),p) = N-1\big) \\ &= \sum_{N=1}^{\infty} \frac{e^{uN}}{N} \binom{N(n-1)}{N-1} p^{N-1} (1-p)^{N(n-1)-(N-1)}. \end{aligned}$$

Our aim is to derive an upper bound for the above expression. Unlike the branching processes considered in Section 3, which were (slightly) supercritical, we are now *subcritical.* Recall that the expected total progeny of a $\mathrm{Bi}(m,p)$ Galton-Watson process is $\frac{1}{1-mp}$; using this fact with $m = n-1$ and $p = \frac{1+\varepsilon}{2(n-1)}$, we see that $\mathbb{E}[Z] = \frac{2}{1-\varepsilon}$.

As $N! \geq (N/e)^N$, we have

$$\frac{1}{N}\binom{N(n-1)}{N-1} \leq \frac{[N(n-1)]^{N-1}}{N!} \leq \frac{(n-1)^{N-1}}{N} e^N, \tag{4.10}$$

which in turn implies that

$$M_Z(u) \leq \sum_{N=1}^{\infty} \frac{e^{uN}}{N}(n-1)^{N-1} e^N \Big(\frac{1+\varepsilon}{2(n-1)}\Big)^{N-1}\Big(1-\frac{1+\varepsilon}{2(n-1)}\Big)^{N(n-1)-(N-1)}.$$

Since $1-x \leq e^{-x}$, we see that

$$M_Z(u) \leq \frac{2}{1+\varepsilon}\sum_{N=1}^{\infty}\frac{1}{N}\Big(\frac{e^{u+1}(1+\varepsilon)e^{-(1+\varepsilon)/2}}{2}\Big)^N e^{\frac{(1+\varepsilon)(N-1)}{2(n-1)}}, \tag{4.11}$$

which is finite for $0 \leq u < \frac{1+\varepsilon}{2} - 1 - \log\frac{1+\varepsilon}{2}$ and n large. Further, it is easily seen that for such u and n the contribution due to terms with $N > C\log n$ is negligible, provided that C is a sufficiently large constant.

Clearly, $C_t(\mathbf{v}_0, i)$ is always bounded above by $\sum_{r=1}^{S_t(i)} Z_r(i)$. In particular, $\sum_{r=1}^{S_T(i)} Z_r(i)$ is an upper bound on $C_T(\mathbf{v}_0, i)$, which equals the number of vertices (i, x) included in the cluster of $\mathbf{v}_0$ from time 0 to time ηV.

Let $u_0 = \frac{1}{2}(\frac{1+\varepsilon}{2} - 1 - \log\frac{1+\varepsilon}{2})$. Since Z is non-negative, $M_Z(u) \geq 1$. It follows from the above (using also the bound $1+x \leq e^x$) that for $0 \leq u \leq u_0$,

$$\begin{aligned} M_{N(\mathbf{v},i)}(u) &\leq M_{S_{\eta V}(i)}(\log M_Z(u)) \leq (1+p(M_Z(u)-1))^{\lceil \eta V\rceil} \\ &\leq \exp\Big(\lceil \eta V\rceil p(M_Z(u)-1)\Big) \\ &= (1+o(1))\exp\Big(\frac{1}{2}\eta n(1+\varepsilon)(M_Z(u)-1)\Big) \\ &\leq (1+o(1))\exp\Big(\Big[\eta n\frac{1+\varepsilon}{1-\varepsilon}u + \frac{1}{2}\eta n(1+\varepsilon)u^2 M_Z''(u_0)\Big]\Big), \end{aligned} \tag{4.12}$$

where the final inequality comes from a second order Taylor expansion. Also we have used the fact that $\mathbb{E}[Z] = 2/(1-\varepsilon)$, and that the second derivative $M_Z''(u)$ is increasing in u for $u \leq u_0$.

Now we run a standard large deviations argument. For all k,

$$\begin{aligned} \mathbb{P}_p(N(\mathbf{v},i) \geq k) &\leq \frac{M_{N(\mathbf{v},i)}(u)}{e^{uk}} \\ &\leq (1+o(1))\exp\Big(\Big(\eta n\frac{1+\varepsilon}{1-\varepsilon} - k\Big)u + \frac{1}{2}\eta n(1+\varepsilon)u^2 M_Z''(u_0)\Big). \end{aligned} \tag{4.13}$$

The expression in (4.13) can be optimised with respect to u in the usual way, and one finds that there exists a constant $c_1 > 0$ such that for all $\nu > 0$,

$$\mathbb{P}_p\Big(N(\mathbf{v},i) \geq (1+\nu)\eta n\frac{1+\varepsilon}{1-\varepsilon}\Big) \leq e^{-c_1\nu\eta n},$$

which yields

$$\mathbb{P}_p\Big(\max_{i=1,\ldots,n} N(\mathbf{v},i) \geq (1+\nu)\eta n\frac{1+\varepsilon}{1-\varepsilon}\Big) \leq n e^{-c_1\nu\eta n}.$$

Since $\varepsilon \leq 1/20$, we have $(1+\varepsilon)/(1-\varepsilon) \leq 11/9$; and therefore there exists a constant $\tilde{c}_1$ such that, for $\nu > 0$,

$$\mathbb{P}_p\big(\max_{i=1,\dots,n} N(\mathbf{v}, i) \geq \frac{11}{9}(1+\nu)\eta n\big) \leq n e^{-2\tilde{c}_1 \nu \eta n}, \tag{4.14}$$

which proves the first statement of Proposition 4.4.

For the remainder, first suppose that $\eta n/\log n \geq 176/\tilde{c}_1$; we may take $\nu = 1/44$ in (4.14) to deduce that

$$\mathbb{P}_p\Big(\max_{i=1,\dots,n} N(\mathbf{v}, i) \geq \frac{5}{4}\eta n\Big) \leq n^{-6} = V^{-3}.$$

Now assume that $\eta n/\log n < 176/\tilde{c}_1$. Note that $N(\mathbf{v}, i)$ is stochastically dominated by $\sum_{r=1}^{\tilde{S}} Z_r(i)$, where $\tilde{S} = \mathrm{Bi}(\lceil \eta V \rceil, p)$. Since $\eta V < 176 \log n n/\tilde{c}_1$, $N(\mathbf{v}, i)$ is stochastically dominated by a $\mathrm{Bi}(\lceil 176\tilde{c}_1^{-1} n \log n \rceil, p)$ random variable. We can perform the moment generating function and large deviations calculations as in (4.11)– (4.13) above, to find that

$$\mathbb{P}_p\Big(\max_{i=1,\dots,n} N(\mathbf{v}, i) \geq \frac{220}{\tilde{c}_1}\log n\Big) \leq n e^{-8\log n(1+o(1))} = o(V^{-3}).$$

Taking $c_2 = 176/\tilde{c}_1$, $c_3 = 220/\tilde{c}_1$ and $c_4 = 1$ completes the proof of Proposition 4.4. ∎

4.3. **Lower bounds on the cluster size and structure.** In this section we establish corresponding lower bounds on the cluster size and structure. We first give a proof of Lemma 4.3, which will in particular establish a lower bound on $\mathbb{P}_p(T = \lceil \eta V \rceil)$, that is, a lower bound on $\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \eta V)$. Our argument will rely on a coupling with a suitable lower bounding Galton-Watson process and the estimates of Proposition 4.4.

Proof of Lemma 4.3. As in Sections 4.1 and 4.2, $C_t(\mathbf{v}_0, i)$ denotes the set of vertices (i, y) (with $i \in \{1, \dots, n\}$ is fixed and $y \in \{1, \dots, n\}$ varying) included in the exploration of cluster $C(\mathbf{v}_0)$ until time t; and $\hat{C}_t(\mathbf{v}_0, i)$ denotes the set of vertices (x, i) included until time t.

Let c_1 be as in Proposition 4.4. Let $m = \frac{5}{4}\max\{\eta n, \frac{176}{c_1}\log n\}$. For each time t, let $\mathcal{E}_t$ be the event that $|C_t(\mathbf{v}_0, i)| \leq m$ and $|\hat{C}_t(\mathbf{v}_0, i)| \leq m$ for all $i = 1, \dots, n$.

Define $\tilde{\Omega} = \Omega - 2m = 2(n-1-m)$. Then, provided that the event $\mathcal{E}_t$ occurs, conditionally on φ_t (that is, given everything else that may have happened until time t), the number of vertices added to $C_t(\mathbf{v}_0)$ as a result of exploring (x_t, y_t) stochastically dominates a $\mathrm{Bi}(\tilde{\Omega}, p)$ random variable. We note that $\Omega - \tilde{\Omega} = O(\eta n + \log n) = o(n)$, since $\eta \to 0$ as $n \to \infty$.

We shall couple our exploration process with a Galton-Watson process starting with a single individual, where the offspring distribution is $\mathrm{Bi}(\tilde{\Omega}, p)$. The mean offspring size for this Galton-Watson process is

$$\tilde{\Omega} p := 1 + \tilde{\varepsilon} = \Big(1 - \frac{m}{n-1}\Big)(1+\varepsilon) = 1 + \varepsilon - O(\eta + n^{-1}\log n) = 1 + \varepsilon(1+o(1)),$$

where we have used the fact that $\eta \ll \varepsilon$ and $n^{-2/3} \ll \varepsilon$. By Proposition 3.3, its survival probability is $2\varepsilon + O(\eta + n^{-1}\log n + \varepsilon^2) = 2\varepsilon(1+o(1))$.

Recall the exploration process and its corresponding colours as described in Section 4.1. Let F_t be the population size of the $\mathrm{Bi}(\tilde{\Omega}, p)$ Galton-Watson process and let F'_t be the set of *green* or active individuals in the Galton-Watson process at time t. Also, let $F = \sup_t F_t$ be the total population size of the Galton-Watson process. Finally, recall that $C_t(\mathbf{v}_0)$ is the set of red and green vertices in the exploration of the cluster of $\mathbf{v}_0$, and $G_t(\mathbf{v}_0)$ the set of green or active vertices in the cluster exploration. By construction, $C_t(\mathbf{v}_0) \subseteq C(\mathbf{v}_0)$ for every $t \geq 0$.

By the above, on the event $\mathcal{E}_t$ intersected with the event that $|C_t(\mathbf{v}_0)| \geq F_t$ and $|G_t(\mathbf{v}_0)| \geq F'_t$, given φ_t, we can couple the Galton-Watson process with the cluster exploration processes for another step so that $|C_{t+1}(\mathbf{v}_0)| \geq F_{t+1}$ and $|G_{t+1}(\mathbf{v}_0)| \geq F'_{t+1}$.

It follows by induction that for each t the random variable $|C_t(\mathbf{v}_0)|I[\mathcal{E}_t]$ is stochastically at least $F_t I[\mathcal{E}_t]$. Hence, for each k,

$$\begin{aligned}\mathbb{P}_p(|C_t(\mathbf{v}_0)| \geq k) &\geq \mathbb{P}_p\big(\mathcal{E}_t \cap \{|C_t(\mathbf{v}_0)| \geq k\}\big) \geq \mathbb{P}_{\Omega,\tilde{\Omega},p}\big(\mathcal{E}_t \cap \{F_t \geq k\}\big)\\ &\geq \mathbb{P}_{\tilde{\Omega},p}(F_t \geq k) - \mathbb{P}_p(\mathcal{E}_t^c),\end{aligned}$$

where $\mathbb{P}_{\Omega,\tilde{\Omega},p}$ denotes the coupling measure. In the second inequality, we have used the fact that for every pair of events $\mathcal{A}, \mathcal{B}$, we have $\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) \geq \mathbb{P}_p(\mathcal{A}) - \mathbb{P}_p(\mathcal{B}^c)$.

By Proposition 4.4, $\mathbb{P}_p(\mathcal{E}_t^c) = O(V^{-3})$, and so, for each $t \leq \eta V$, we obtain

$$\mathbb{P}_p(|C_t(\mathbf{v}_0)| \geq k) \geq \mathbb{P}_{\tilde{\Omega},p}(F_t \geq k) + O(V^{-3}),$$

which establishes (4.4). Similarly, for each time t and non-negative integer k,

$$\mathbb{P}_p(|G_t(\mathbf{v}_0)| \geq k) \geq \mathbb{P}_{\tilde{\Omega},p}(F'_t \geq k) - \mathbb{P}_p(\mathcal{E}_t^c),$$

and, in particular, for each $t \leq \eta V$,

$$\mathbb{P}_p(T \geq t) = \mathbb{P}_p(|G_t(\mathbf{v}_0)| \geq 1) \geq \mathbb{P}_{\tilde{\Omega},p}(F'_t \geq 1) + O(V^{-3}).$$

Notice that, for $t \leq \eta V$,

$$\mathbb{P}_{\tilde{\Omega},p}(F'_t = 0) \leq \mathbb{P}_{\tilde{\Omega},p}(F'_{\lceil \eta V \rceil} = 0) \leq \mathbb{P}_{\tilde{\Omega},p}(F < \infty).$$

In this way we arrive at

$$\begin{aligned}\mathbb{P}_p(|C_t(\mathbf{v}_0)| \geq k) &\geq \mathbb{P}_{\tilde{\Omega},p}(F_t \geq k) + O(V^{-3})\\ &\geq \mathbb{P}_{\tilde{\Omega},p}(F_t \geq k, F'_t > 0) + O(V^{-3})\\ &= \mathbb{P}_{\tilde{\Omega},p}(F'_t > 0) - \mathbb{P}_{\tilde{\Omega},p}(F'_t > 0, F_t < k) + O(V^{-3})\\ &\geq \mathbb{P}_{\tilde{\Omega},p}(F'_t > 0) - \mathbb{P}_p(\mathrm{Bi}(t\tilde{\Omega}, p) < k) + O(V^{-3})\\ &\geq \mathbb{P}_{\tilde{\Omega},p}(F = \infty) - \mathbb{P}_p(\mathrm{Bi}(t\tilde{\Omega}, p) < k) + O(V^{-3}),\end{aligned}$$

since on the event that the process is alive at time t and the event $\mathcal{E}_t$ occurs, we can couple the number of vertices added at all steps until t so that it is at least as large as a sum of t independent binomials $\mathrm{Bi}(\tilde{\Omega}, p)$.

Hence, also using the facts that $p \geq 1/\Omega$ and $|\Omega - \tilde{\Omega}| = o(\Omega)$, for every constant $\delta \in (0,1)$ there is a constant $\alpha > 0$ such that

$$\begin{aligned}\mathbb{P}_p\Big(|C(\mathbf{v}_0)| \geq (1-\delta)\eta V\Big) &\geq \mathbb{P}_p(|C_{\eta V}(\mathbf{v}_0)| \geq (1-\delta)\eta V) \qquad (4.15)\\ &\geq \mathbb{P}_{\tilde{\Omega},p}(F = \infty) - \mathbb{P}_p\big(\mathrm{Bi}(\eta V\tilde{\Omega}, p) < (1-\delta)\eta V\big) + O(V^{-3})\\ &= 2\varepsilon + O(\eta + n^{-1}\log n + \varepsilon^2) + e^{-\alpha\eta V} + O(V^{-3}).\end{aligned}$$

But equally, we could run the exploration process until time $(1+\delta)\eta V$ to obtain a cluster of size ηV **whp**, that is, we could use the above with η replaced by $\eta/(1-\delta)$ to obtain that

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \eta V) \geq 2\varepsilon + O(\eta + n^{-1}\log n + \varepsilon^2) + e^{-\alpha\eta V} + O(V^{-3}). \qquad (4.16)$$

This establishes (4.5), thus completing the proof of Lemma 4.3, and hence also the proof of Proposition 2.1. ∎

Let us call a horizontal line *good* if it contains at least $\eta V/(4n) = \eta n/4$ elements in $C(\mathbf{v_0})$ along that line, and *bad* otherwise. We shall now prove Proposition 2.4, thus establishing a lower bound on the number of good lines.

Proof of Proposition 2.4. As earlier, for a vertex $\mathbf{v}_0$ and $i \in \{1, \ldots, n\}$, the random variable $C_t(\mathbf{v}_0, i)$ denotes the number of elements of the i-th horizontal line contained in $C_t(\mathbf{v}_0)$, the part of $C(\mathbf{v}_0)$ obtained by running the exploration process until time t. Also, $C(\mathbf{v}_0, i)$ is the number of elements of the i-th horizontal line in $C(\mathbf{v}_0)$ and $N(\mathbf{v}_0, i)$ is the number of such elements included in $C(\mathbf{v}_0)$ until time $\lceil \eta V \rceil$.

Let c_2 be as in Proposition 4.4, statement (1), and choose $C = c_2$. By Proposition 4.4, statement (1), we have

$$\mathbb{P}_p\Big(\max_{\mathbf{v}_0} \max_{i=1,\ldots,n} : N(\mathbf{v}_0, i) \leq \frac{5}{4}\eta n \Big) = O(V^{-3}).$$

We select a vertex $\mathbf{v}_0$. Let $\mathcal{A}$ be the event that $\{\max_{i=1,\ldots,n} N(\mathbf{v}_0, i) \leq \frac{5}{4}\eta n\}$. Let also $\mathcal{B}$ ("$\mathcal{B}$" for "bad") be the event that fewer than $3n/4$ lines are good for the cluster $C(\mathbf{v}_0)$. On the event that $|C(\mathbf{v}_0)| \geq \eta V$, we have $|C_{\lceil \eta V \rceil}(\mathbf{v}_0)| \geq |R_{\lceil \eta V \rceil}(\mathbf{v}_0)| = \lceil \eta V \rceil$. It follows that we only need to show that

$$\mathbb{P}_p\left(\mathcal{B} \cap \{|C(\mathbf{v}_0)| \geq \eta V\}\right) = \mathbb{P}_p\left(\mathcal{B} \cap \{|C_{\lceil \eta V \rceil}(\mathbf{v}_0)| \geq \eta V\}\right) = o(V^{-1}). \tag{4.17}$$

Indeed, summing over all vertices $\mathbf{v}_0$ we may deduce from (4.17) that **whp** there is no $\mathbf{v}_0$ such that $|C(\mathbf{v}_0)| \geq \eta V$ and fewer than $3n/4$ lines are good for $C(\mathbf{v}_0)$. In order to establish (4.17), we shall show that

$$\mathbb{P}_p(\mathcal{B} \cap \{|C_{\lceil \eta V \rceil}(\mathbf{v}_0)| \geq \eta V\}) \leq \mathbb{P}_p(\mathcal{A}^c). \tag{4.18}$$

Clearly, $|C(\mathbf{v}_0, i)| \geq N(\mathbf{v}_0, i)$ for every i. Let us write $g_{\mathbf{v}_0}$ and $b_{\mathbf{v}_0}$ respectively for the number of good and bad lines in $C_{\lceil \eta V \rceil}(\mathbf{v}_0)$.

On the event $\mathcal{A}$, the explored cluster $C_{\lceil \eta V \rceil}(\mathbf{v}_0)$ at time ηV contains at most $5\eta n/4$ elements of every good line and at the same time has size at least ηV. Hence, using also that $g_{\mathbf{v}} = n - b_{\mathbf{v}_0}$, on $\mathcal{A} \cap \{|C_{\lceil \eta V \rceil}(\mathbf{v}_0)| \geq \eta V\}$,

$$\eta V \leq |C_{\eta V}(\mathbf{v}_0)| \leq \frac{5}{4}\eta n g_{\mathbf{v}_0} + \frac{1}{4}\eta n b_{\mathbf{v}_0} = \eta n g_{\mathbf{v}_0} + \frac{1}{4}\eta V,$$

which gives

$$\frac{3}{4}\eta V \leq \eta n g_{\mathbf{v}_0}$$

and hence

$$g_{\mathbf{v}_0} \geq \frac{3}{4} n.$$

In other words, on $\mathcal{A} \cap \{|C_{\lceil \eta V \rceil}(\mathbf{v}_0)| \geq \eta V\}$, the number of good lines is at least $3n/4$, which means that

$$\mathbb{P}_p(\mathcal{B} \cap \mathcal{A} \cap \{|C_{\lceil \eta V \rceil}(\mathbf{v}_0)| \geq \eta V\}) = 0, \tag{4.19}$$

and so establishes claim (4.18). Then, from Proposition 4.4, we see that

$$\mathbb{P}_p(\mathcal{B} \cap \{|C_{\lceil \eta V \rceil}(\mathbf{v}_0)| \geq \eta V\}) \leq \mathbb{P}_p(\mathcal{A}^c) = O(V^{-3}), \tag{4.20}$$

as required. This completes the proof of Proposition 2.4. ∎

5. Concentration of measure for the number of vertices in large clusters

This section contains our proof of Proposition 2.2. In outline, the goal is to establish concentration of measure for $Z_{\geq \alpha^{-2}}$, for an appropriate choice of $\alpha \ll \varepsilon$ to be determined below. This will be carried out by second moment methods, in a slightly unusual way, as we explain now.

For every ℓ, define the centered versions of the random variables $Z_{\geq \ell}$ by

$$\bar{Z}_{\geq \ell} = Z_{\geq \ell} - \mathbb{E}_p[Z_{\geq \ell}]. \tag{5.1}$$

The entire proof revolves around two scales of magnitude, denoted by $\underline{N}$ and $\overline{N}$. The value $\overline{N}$ is the large scale, and corresponds to ε_0^{-2} in Proposition 2.2. The value $\underline{N}$ is the smaller scale, and will be determined below. The scales $\underline{N}$ and $\overline{N}$ are related through a positive integer I defined by

$$\overline{N} = \underline{N} 2^I. \tag{5.2}$$

With this notation, proving Proposition 2.2 amounts to establishing that

$$\mathbb{P}_p\Big(|\bar{Z}_{\geq \overline{N}}| \geq \delta\varepsilon V\Big) = o(1). \tag{5.3}$$

We first observe that

$$|\bar{Z}_{\geq \overline{N}}| \leq |\bar{Z}_{\geq \underline{N}}| + \sum_{i=0}^{I-1} |\bar{Z}_{\geq 2^{i+1}\underline{N}} - \bar{Z}_{\geq 2^{i}\underline{N}}|.$$

The goal is now to establish sufficient bounds on the variances of the above random variables, so that we can prove that $\bar{Z}_{\geq \overline{N}}$ is concentrated. For this, we choose a sequence $\{\delta_i\}_{i=0}^{I-1}$ such that each $\delta_i > 0$ and $\sum_{i=0}^{I-1} \delta_i \leq \frac{\delta}{2}$. If $|\bar{Z}_{\geq \overline{N}}| \geq \delta\varepsilon V$, then either $|\bar{Z}_{\geq \underline{N}}| \geq \delta\varepsilon V/2$, or $|\bar{Z}_{\geq 2^{i+1}\underline{N}} - \bar{Z}_{\geq 2^{i}\underline{N}}| \geq \delta_i\varepsilon V$ for some $0 \leq i \leq I-1$. Consequently,

$$\mathbb{P}_p\Big(|\bar{Z}_{\geq \overline{N}}| \geq \delta\varepsilon V\Big) \leq \mathbb{P}_p\big(|\bar{Z}_{\geq \underline{N}}| \geq \delta\varepsilon V/2\big) + \sum_{i=0}^{I-1} \mathbb{P}_p\Big(|\bar{Z}_{\geq 2^{i+1}\underline{N}} - \bar{Z}_{\geq 2^{i}\underline{N}}| \geq \delta_i\varepsilon V\Big), \tag{5.4}$$

and we are going to upper bound each term on the right hand side separately. Our argument relies on estimating the variance of $Z_{\geq \underline{N}}$ and those of the differences $Z_{\geq 2^{i+1}\underline{N}} - Z_{\geq 2^{i}\underline{N}}$. This is accomplished in Section 5.2 – see Lemmas 5.2 and Lemma 5.3. The variance estimates impose various restrictions on $\underline{N}$ and $\overline{N}$; in Section 5.3 we show that these can be satisfied as long as $\varepsilon^3 V \gg \log n$, which establishes Proposition 2.2. The key to the proof is to choose $\underline{N}$, $\overline{N}$ and $\{\delta_i\}_{i=0}^{I-1}$ so as to ensure adequate concentration of measure.

The remainder of this section is organised as follows. In Section 5.1 we bound the cluster tail bounds of the form $\mathbb{P}_p(|C(\mathbf{v}_0)| \in [\ell, 2\ell])$; these are needed to estimate the distribution of the random variables $Z_{\geq 2^{i+1}\underline{N}} - Z_{\geq 2^{i}\underline{N}}$. Here we shall make use of Galton-Watson processes estimates and comparisons established in Sections 3–4. Then, in Section 5.2, we upper bound the variances of $Z_{\geq \underline{N}}$ and $Z_{\geq 2^{i+1}\underline{N}} - Z_{\geq 2^{i}\underline{N}}$. Finally, in Section 5.3 we complete our proof of Proposition 2.2.

5.1. **Key ingredients.** As before, for a positive integer N and an edge probability p, $\mathbb{P}_{N,p}$ denotes the probability measure corresponding to the Galton-Watson process where the family size is a $\mathrm{Bi}(N,p)$ random variable; also, F is the total progeny.

The remainder of this section is devoted to establishing a bound on the cluster tail crucial to the arguments in Sections 5.2 and Section 5.3. Recall that $\Omega = 2(n-1)$, choose a positive integer ℓ, and suppose that $\tilde{\Omega} = \tilde{\Omega}(n)$ satisfies $\Omega - \tilde{\Omega} = O(\log n + \ell/n)$ for some $\ell = o(\varepsilon V)$. Suppose further that $\varepsilon = \varepsilon(n) = \Omega p - 1 \to 0$ such that $V^{-1/3} \ll \varepsilon \ll 1$ for $V = n^2$ sufficiently large. Then

$|\varepsilon - \tilde{\varepsilon}| = p|\Omega - \tilde{\Omega}| = O(\log n/n + \ell/n^2) = o(\varepsilon)$, since $\ell = o(\varepsilon V)$, and so we may use the results of Proposition 3.1. Hence, as long as $\ell = o(\varepsilon V)$, we have, uniformly in n,

$$|\mathbb{P}_{\Omega,p}(F \geq \ell) - \mathbb{P}_{\tilde{\Omega},p}(F \geq \ell)| \leq C\Big(p|\Omega - \tilde{\Omega}| + \frac{1}{n\ell^{1/2}} + \frac{1}{\ell^3}\Big). \tag{5.5}$$

We shall use inequality (5.5) in the following lemma to identify the cluster tail distribution more precisely.

Lemma 5.1 (Bound on the cluster tail)**.** *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Let $V^{-1/3} \ll \varepsilon \ll 1$, and let $\ell \in \mathbb{N}$ satisfy $\ell \leq V^{2/3}$ and $\ell \ll \varepsilon V$. Then there exists a constant C such that, for n sufficiently large,*

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \in [\ell, 2\ell]) \leq \frac{C}{\sqrt{\ell}}. \tag{5.6}$$

Proof. By Lemma 4.1,

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) \leq \mathbb{P}_{\Omega,p}(F \geq \ell).$$

Further, by Lemma 4.3, C can be chosen large enough that

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq 2\ell) \geq \mathbb{P}_{\tilde{\Omega},p}(F \geq 2\ell) + O(V^{-3}),$$

where $\tilde{\Omega} = \Omega - \frac{5}{2}\max\{\ell n^{-1}, C\log n\}$. Let $\tilde{\varepsilon} = \tilde{\Omega}p - 1$ and note that $|\varepsilon - \tilde{\varepsilon}| \leq \frac{C}{n^2}(\ell + n\log n)$ (after a suitable adjustment of C). Since $\varepsilon \geq V^{-1/3} = n^{-2/3}$, our assumptions on ℓ imply that $|\varepsilon - \tilde{\varepsilon}| = o(\varepsilon)$. By Propositions 3.1 and 3.2,

$$\begin{aligned}\mathbb{P}_{\tilde{\Omega},p}(F \geq 2\ell) &\geq \mathbb{P}_{\Omega,p}(F \geq 2\ell) + O\big(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{n\ell^{1/2}} + \frac{1}{\ell^3}\big)\\ &\geq \mathbb{P}_{\Omega,p}(F \geq \ell) + O(|\varepsilon - \tilde{\varepsilon}|) + O(\ell^{-1/2}).\end{aligned}$$

It follows that

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \geq \ell) - \mathbb{P}_p(|C(\mathbf{v}_0)| \geq 2\ell) \leq O(|\varepsilon - \tilde{\varepsilon}|) + O(\ell^{-1/2}),$$

and we only need to show that $|\varepsilon - \tilde{\varepsilon}|$ is $O(l^{-1/2})$. This is equivalent to showing that both $\frac{\ell}{n^2}$ and $\frac{\log n}{n}$ are $O(l^{-1/2})$. The condition $\ell \leq V^{2/3}$ is equivalent to $\frac{\ell}{n^2} \leq \ell^{-1/2}$. The bound $\frac{\log n}{n} \leq \ell^{-1/2}$ holds when $\ell \leq n^2/(\log n)^2$; as $\ell \leq V^{2/3} = n^{4/3}$, this is also true for n sufficiently large. ∎

5.2. **Variance estimates.** Lemmas 5.2 and 5.3 below contain variance estimates essential to our proof of Proposition 2.2.

Lemma 5.2 (Variance of the number of vertices in moderate clusters)**.** *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Suppose that $V^{-1/3} \ll \varepsilon \ll 1$. Choose N such that $N = o(\sqrt{V})$ and $N = o(\varepsilon^2 V)$. Then*

$$\mathrm{Var}_p(Z_{\geq N}) = o\big((\varepsilon V)^2\big). \tag{5.7}$$

Proof. First note that $\mathrm{Var}_p(Z_{\geq N}) = \mathrm{Var}_p(Z_{<N})$, where

$$Z_{<N} = V - Z_{\geq N} = \sum_{\mathbf{v}} I[|C(\mathbf{v})| < N].$$

We expand $\mathrm{Var}_p(Z_{<N})$ as

$$\mathrm{Var}_p(Z_{<N}) = \sum_{\mathbf{v}_0,\mathbf{v}_1} \big[\mathbb{P}_p\big(|C(\mathbf{v}_0)| < N, |C(\mathbf{v}_1)| < N\big) - \mathbb{P}_p(|C(\mathbf{v}_0)| < N)^2\big]. \tag{5.8}$$

We separate each term involving distinct $\mathbf{v_0}$ and $\mathbf{v_1}$ into two, according to whether or not $\mathbf{v}_1 \in C(\mathbf{v}_0)$. We can then write

$$\mathrm{Var}_p(Z_{<N}) = S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} + S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1}, \tag{5.9}$$

where $S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1}(N)$, $S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} = S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1}(N)$ and

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0, \mathbf{v}_1} \mathbb{P}_p(|C(\mathbf{v}_0)| < N, \mathbf{v}_1 \in C(\mathbf{v}_0)), \tag{5.10}$$

$$S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0, \mathbf{v}_1} \Big[\mathbb{P}_p(|C(\mathbf{v}_0)| < N, |C(\mathbf{v}_1)| < N, \mathbf{v}_1 \notin C(\mathbf{v}_0)) - \mathbb{P}_p(|C(\mathbf{v}_0)| < N)^2\Big]. \tag{5.11}$$

It is easily seen that

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = V\mathbb{E}_p\big[|C(\mathbf{v}_0)| I[|C(\mathbf{v}_0)| < N]\big], \tag{5.12}$$

and we upper bound

$$\begin{aligned}\mathbb{E}_p\big[|C(\mathbf{v}_0)| I[|C(\mathbf{v}_0)| < N]\big] &= \sum_{l=1}^{N} \mathbb{P}_p(l \le |C(\mathbf{v}_0)| < N) \le \sum_{l=1}^{N} \mathbb{P}_p(|C(\mathbf{v}_0)| \ge l) \\ &\le C \sum_{l=1}^{N} \big(\varepsilon + \frac{1}{\sqrt{l}}\big),\end{aligned} \tag{5.13}$$

where the last inequality follows from Lemma 4.2. It follows that

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = O(VN\varepsilon + V\sqrt{N}) = o(\varepsilon^2 V^2), \tag{5.14}$$

provided $N = o(\varepsilon V)$ and $N = o(\varepsilon^4 V^2)$. When $\varepsilon^3 V \gg 1$ then $\varepsilon V \ll \varepsilon^4 V^2$, so only the first constraint on N is binding, i.e. $S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = o(\varepsilon^2 V^2)$ as long as $N = o(\varepsilon V)$.

To upper bound $S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1}$ note that, by [4, inequality (9.7)],

$$S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} \le p \sum_{\{\mathbf{u}, \mathbf{v}\}} \mathbb{E}_p\Big[|C(\mathbf{u})||C(\mathbf{v})| I[|C(\mathbf{u})| < N, |C(\mathbf{v})| < N, \mathbf{v} \notin C(\mathbf{u})]\Big], \tag{5.15}$$

where the summation is over all edges $\{\mathbf{u}, \mathbf{v}\}$ of $H(2, n)$. We can estimate this similarly to $S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1}$ above, and find that

$$\begin{aligned}S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} &\le p \sum_{(\mathbf{u}, \mathbf{v})} \sum_{l_1, l_2 = 1}^{N} \mathbb{P}_p\Big(l_1 \le |C(\mathbf{u})| < N, l_2 \le |C(\mathbf{v})| < N, \mathbf{v} \notin C(\mathbf{u})\Big) \\ &\le p \sum_{(\mathbf{u}, \mathbf{v})} \sum_{l_1, l_2 = 1}^{N} \mathbb{P}_p\Big(|C(\mathbf{u})| \ge l_1, |C(\mathbf{v})| \ge l_2, \mathbf{v} \notin C(\mathbf{u})\Big).\end{aligned} \tag{5.16}$$

Since $\mathbf{v} \notin C(\mathbf{u})$, $|C(\mathbf{u})|$ and $|C(\mathbf{v})|$ are each *independently* of one another stochastically dominated by the total progeny of a $\mathrm{Bi}(\Omega, p)$ Galton-Watson process. (To see this in more detail, think of first constructing the cluster of $\mathbf{u}$, and subsequently construct the cluster of $\mathbf{v}$ in the smaller graph with $C(\mathbf{u})$ removed.) Using Lemma 4.2, we then see that, since Ωp is bounded above as $n \to \infty$,

$$\begin{aligned}S_{\mathbf{v}_0 \not\leftrightarrow \mathbf{v}_1} &\le C\Omega p V \sum_{l_1, l_2 = 1}^{N} \Big(\varepsilon + \frac{1}{\sqrt{l_1}}\Big)\Big(\varepsilon + \frac{1}{\sqrt{l_2}}\Big) \\ &\le CV\big(\varepsilon N + \sqrt{N}\big)^2 \le O(V\varepsilon^2 N^2 + VN).\end{aligned} \tag{5.17}$$

Thus, as long as $N = o(\sqrt{V})$ and $N = o(\varepsilon^2 V)$,

$$S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1} = o\big((\varepsilon V)^2\big), \tag{5.18}$$

which completes the proof. ∎

Lemma 5.3 (Variance of the number of vertices in intermediate clusters)**.** *Set $p = p_c + \frac{\varepsilon}{\Omega}$. Assume that $V^{-1/3} \ll \varepsilon \ll 1$, and take $N \in \mathbb{N}$ such that $N \leq V^{2/3}$ and $N \ll \varepsilon V$. Then*

$$\mathrm{Var}_p(Z_{\geq N} - Z_{\geq 2N}) \leq \frac{CV^2}{\sqrt{N}}\Big(\frac{\log n}{n} \bigvee \frac{N}{V} \bigvee \frac{1}{N^3}\Big). \tag{5.19}$$

Proof. We have

$$\mathrm{Var}_p(Z_{\geq N} - Z_{\geq 2N}) = \sum_{\mathbf{v}_0, \mathbf{v}_1} \mathbb{P}_p(|C(\mathbf{v}_0)| \in (N, 2N], |C(\mathbf{v}_1)| \in (N, 2N]) - \mathbb{P}_p(|C(\mathbf{v}_0)| \in (N, 2N])^2.$$

Once again we split the sum according to whether or not $\mathbf{v}_1 \in C(\mathbf{v}_0)$, obtaining

$$\mathrm{Var}_p(Z_{\geq N} - Z_{\geq 2N}) = S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} + S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1}, \tag{5.20}$$

where now

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0, \mathbf{v}_1} \mathbb{P}_p\Big(|C(\mathbf{v}_0)| \in (N, 2N], \mathbf{v}_1 \in C(\mathbf{v}_0)\Big), \tag{5.21}$$

$$S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0, \mathbf{v}_1} \Big[\mathbb{P}_p\Big(|C(\mathbf{v}_0)|, |C(\mathbf{v}_1)| \in (N, 2N], \mathbf{v}_1 \notin C(\mathbf{v}_0)\Big) - \mathbb{P}_p\Big(|C(\mathbf{v}_0)| \in (N, 2N]\Big)^2\Big]. \tag{5.22}$$

Just as in the proof of Lemma 5.2,

$$S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1} = V\mathbb{E}_p\big[|C(\mathbf{v}_0)| I[N < |C(\mathbf{v}_0)| \leq 2N]\big] \leq CV\sqrt{N}. \tag{5.23}$$

But

$$V\sqrt{N} \leq \frac{V^2}{\sqrt{N}}\Big(\frac{\log n}{n} \bigvee \frac{N}{V}\Big), \tag{5.24}$$

and so $S_{\mathbf{v}_0 \leftrightarrow \mathbf{v}_1}$ is bounded by the right hand side of (5.19).

Dealing with $S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1}$ requires more effort. Define

$$p_{\mathbf{v}_0, \mathbf{v}_1} := \mathbb{P}_p\Big(|C(\mathbf{v}_0)| \in (N, 2N], |C(\mathbf{v}_1)| \in (N, 2N], \mathbf{v}_1 \notin C(\mathbf{v}_0)\Big) - \mathbb{P}_p\big(|C(\mathbf{v}_0)| \in (N, 2N]\big)^2, \tag{5.25}$$

so that

$$S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0, \mathbf{v}_1} p_{\mathbf{v}_0, \mathbf{v}_1}.$$

Now rewrite

$$\frac{p_{\mathbf{v}_0, \mathbf{v}_1}}{\mathbb{P}_p(|C(\mathbf{v}_0)| \in (N, 2N])} = \mathbb{P}_p\Big(|C(\mathbf{v}_1)| \in (N, 2N], \mathbf{v}_1 \notin C(\mathbf{v}_0) \big| |C(\mathbf{v}_0)| \in (N, 2N]\Big) - \mathbb{P}_p\big(|C(\mathbf{v}_1)| \in (N, 2N]\big).$$

Recall that $N(\mathbf{v}, i)$ is the number of elements in the i-th horizontal line included in the cluster until time ηn^2. The proof of Proposition 4.4 implies that there is some constant $C > 0$ such that **wvhp** every $\mathbf{v}$ such that $|C(\mathbf{v})| \in (N, 2N]$ satisfies

$$N(\mathbf{v}, i) \leq C\Big[\log n \vee \frac{N}{n}\Big]. \tag{5.26}$$

(To see this, think of running the exploration process with stopping time T as in (4.1), where η is defined by $2N = \lceil \eta V \rceil$ (so in particular $\eta \ll \varepsilon$). Since $|C(\mathbf{v})| \in (N, 2N]$, we have $C(\mathbf{v}) = C_T(\mathbf{v})$.) Letting

$$\tilde{\Omega} = \Omega - 2C\Big[\log n \vee \frac{N}{n}\Big], \tag{5.27}$$

we can lower bound

$$\mathbb{P}_p(|C(\mathbf{v}_0)| \in (N, 2N]) \geq \mathbb{P}_{\tilde{\Omega},p}(F \geq N) - \mathbb{P}_{\Omega,p}(F \geq 2N) + O(V^{-3}). \tag{5.28}$$

Further, we can upper bound

$$\begin{aligned}
&\mathbb{P}_p(|C(\mathbf{v}_1)| \in (N, 2N], \mathbf{v}_1 \notin C(\mathbf{v}_0) \big| |C(\mathbf{v}_0)| \in (N, 2N]) \\
&\quad = \mathbb{P}_p(|C(\mathbf{v}_1)| \geq N, \mathbf{v}_1 \notin C(\mathbf{v}_0) \big| |C(\mathbf{v}_0)| \in (N, 2N]) \\
&\qquad - \mathbb{P}_p(|C(\mathbf{v}_1)| \geq 2N, \mathbf{v}_1 \notin C(\mathbf{v}_0) \big| |C(\mathbf{v}_0)| \in (N, 2N]) \\
&\quad \leq \mathbb{P}_{\Omega,p}(F \geq N) - \mathbb{P}_{\tilde{\Omega},p}(F \geq 2N) + O(V^{-3}).
\end{aligned} \tag{5.29}$$

(Once again, to see this, think of first exploring the cluster of $\mathbf{v}_0$ and, after that, the cluster of $\mathbf{v}_1$ in $H(2, n)$ with the cluster of $\mathbf{v}_0$ removed.)

Since $N \leq V^{2/3}$ and $N \ll \varepsilon V$, we can use Lemma 5.1 to bound $\mathbb{P}_p\big(|C(\mathbf{v}_0)| \in (N, 2N]\big)$ and obtain

$$p_{\mathbf{v}_0,\mathbf{v}_1} \leq \frac{C}{\sqrt{N}}\Big(\mathbb{P}_{\Omega,p}(F \geq N) - \mathbb{P}_{\tilde{\Omega},p}(F \geq N) + \mathbb{P}_{\Omega,p}(F \geq 2N) - \mathbb{P}_{\tilde{\Omega},p}(F \geq 2N)\Big) + O(V^{-3}). \tag{5.30}$$

By (5.5), with $\varepsilon = p\Omega - 1$ and $\tilde{\varepsilon} = p\tilde{\Omega} - 1$, for every $\ell \in \mathbb{N}$,

$$\mathbb{P}_{\Omega,p}(F \geq \ell) - \mathbb{P}_{\tilde{\Omega},p}(F \geq \ell) \leq C\Big(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{n\ell^{1/2}} + \frac{1}{\ell^3}\Big). \tag{5.31}$$

Note that, by (5.27),

$$|\varepsilon - \tilde{\varepsilon}| = C\Big[\frac{\log n}{n} \bigvee \frac{N}{n^2}\Big], \tag{5.32}$$

so that we always have $\frac{1}{n\ell^{1/2}} = O(|\varepsilon - \tilde{\varepsilon}|)$.

Consequently (with the value of C adjusted between inequalities), for all vertex pairs $\mathbf{v}_0, \mathbf{v}_1$,

$$p_{\mathbf{v}_0,\mathbf{v}_1} \leq \frac{C}{\sqrt{N}}\Big(|\varepsilon - \tilde{\varepsilon}| + \frac{1}{N^3}\Big). \tag{5.33}$$

Summing over $\mathbf{v}_0, \mathbf{v}_1$,

$$S_{\mathbf{v}_0 \nleftrightarrow \mathbf{v}_1} = \sum_{\mathbf{v}_0,\mathbf{v}_1} p_{\mathbf{v}_0,\mathbf{v}_1} \leq \frac{CV^2}{\sqrt{N}}\Big(|\varepsilon - \tilde{\varepsilon}| \vee \frac{1}{N^3}\Big) = \frac{CV^2}{\sqrt{N}}\Big(\frac{\log n}{n} \bigvee \frac{N}{V} \bigvee \frac{1}{N^3}\Big), \tag{5.34}$$

since $|\varepsilon - \tilde{\varepsilon}| = O\Big(\frac{\log n}{n} \vee \frac{N}{n^2}\Big)$. ∎

5.3. **Proof of Proposition 2.2.** We are now ready to complete the proof of Proposition 2.2. We will make essential use of Lemmas 5.2 and 5.3. The choice ε_0 in Proposition 2.2 will be given by $\varepsilon_0^{-2} = \overline{N}$, where $\overline{N}$ is determined below.

Let $\delta > 0$, and, for $i = 0, 1, \ldots, I - 1$, let

$$\delta_i = \frac{\delta}{4\zeta(2)[(i+1) \wedge (I-i)]^2}. \tag{5.35}$$

The reasons for our choice for $\{\delta_i\}_{i=0}^{I-1}$ will become apparent shortly. For now let us note that

$$\sum_{i=1}^{I-1} \delta_i \leq \delta/2. \tag{5.36}$$

Recall the definition of $\bar{Z}_{\geq \ell}$ from (5.1) and the decomposition in (5.4). We will prove that the right hand side of (5.4) is $o(1)$ for suitable $\underline{N}$ and $\overline{N}$; the conditions that $\underline{N}$ and $\overline{N}$ must satisfy are as follows:

$$\underline{N} \ll n, \qquad \underline{N} \ll \varepsilon^2 V, \qquad \underline{N} \gg \frac{(\log n)^2}{n^2\varepsilon^4}, \qquad \underline{N} \geq \Big(\frac{n}{\log n}\Big)^{1/3}, \tag{5.37}$$

$$\overline{N} \ll \varepsilon V, \qquad \overline{N} \leq V^{2/3}. \tag{5.38}$$

Finally, Proposition 2.2 requires that $\overline{N} \gg \varepsilon^{-2}$. As $\varepsilon \gg (\log V)^{1/3} V^{-1/3} = (\log V)^{1/3} n^{-2/3}$, the choices $\underline{N} = n\varepsilon^{1/2}(\log V)^{1/2}$ and $\overline{N} = V^{2/3} = n^{4/3}$ clearly satisfy the bounds in (5.37)–(5.38); thus we have proved that appropriate choices can be made.

Let us note that it is here that the condition $\varepsilon \gg (\log n)^{1/3} V^{-1/3}$ in Theorem 1.1 arises. We need to show concentration of measure for clusters of size $\underline{N}$, which satisfies the constraint $\underline{N} \ll n$; for such clusters , we are unable to control very precisely the number of vertices per coordinate line (see Proposition 4.4) – this then gives rise to the $\log n/n$ factor in Lemma 5.3, and hence at this point in our proof.

We now prove that the concentration bound in Proposition 2.2 holds. By (5.37), $\underline{N}$ satisfies the hypotheses of Lemma 5.2; hence, using the Chebyshev inequality,

$$\mathbb{P}_p\big(|\bar{Z}_{\geq \underline{N}}| \geq \delta\varepsilon V/2\big) \leq \frac{4\mathrm{Var}_p(Z_{\geq \underline{N}})}{(\delta\varepsilon V)^2} = o(1). \tag{5.39}$$

Denote $N_i = 2^{i+1}\underline{N}$, and recall the relation between $\underline{N}$ and $\overline{N}$ in (5.2). Since $N_i \leq \overline{N}$, (5.38) implies that $N_i \ll \varepsilon V$ and $N_i \leq V^{2/3}$ for each i. Therefore, applying Lemma 5.3 to $N_i = 2^{i+1}\underline{N}$ and using the Chebyshev inequality, we obtain

$$\begin{aligned}\mathbb{P}_p\Big(|\bar{Z}_{\geq 2^{i+1}\underline{N}} - \bar{Z}_{\geq 2^i \underline{N}}| \geq \delta_i\varepsilon V\Big) &\leq \big(\delta_i\varepsilon V\big)^{-2}\mathrm{Var}_p\big(Z_{\geq 2^{i+1}\underline{N}} - Z_{\geq 2^i\underline{N}}\big)\\ &\leq \big(\delta_i\varepsilon V\big)^{-2}\Big[\frac{CV^2}{\sqrt{N_i}}\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V} \bigvee \frac{1}{N_i^3}\Big)\Big].\end{aligned} \tag{5.40}$$

It follows that under our assumptions

$$\mathbb{P}_p\big(|\bar{Z}_{\geq \overline{N}}| \geq \delta\varepsilon V\big) \leq o(1) + \sum_{i=0}^{I-1} \frac{\frac{CV^2}{\sqrt{N_i}}\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V} \bigvee \frac{1}{N_i^3}\Big)}{(\delta_i\varepsilon V)^2}. \tag{5.41}$$

Each term here is given by

$$\frac{CV^2}{\sqrt{N_i}} \frac{\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V} \bigvee \frac{1}{N_i^3}\Big)}{(\delta_i\varepsilon V)^2} = \frac{C}{\sqrt{N_i}} \frac{\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V} \bigvee \frac{1}{N_i^3}\Big)}{\delta_i^2\varepsilon^2}. \tag{5.42}$$

By the last assumption in (5.37), for all i, $\frac{\log n}{n} \geq \frac{1}{\underline{N}^3} \geq \frac{1}{N_i^3}$, so that the last term is never equal to the maximum. It follows that we need to upper bound

$$\mathbb{P}_p\big(|\bar{Z}_{\geq \overline{N}}| \geq \delta\varepsilon V\big) \leq o(1) + \sum_{i=0}^{I-1} \frac{\frac{CV^2}{\sqrt{N_i}}\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V}\Big)}{(\delta_i\varepsilon V)^2}. \tag{5.43}$$

Letting m be the smallest i such that

$$\frac{\log n}{n} \leq \frac{N_i}{V}, \tag{5.44}$$

we can write

$$\sum_{i=0}^{I-1} \frac{\frac{CV^2}{\sqrt{N_i}}\Big(\frac{\log n}{n} \bigvee \frac{N_i}{V}\Big)}{(\delta_i \varepsilon V)^2} = \sum_{i=0}^{m} \frac{C}{\sqrt{N_i}} \frac{\log n}{n\delta_i^2\varepsilon^2} + \sum_{i=m+1}^{I-1} C\frac{\sqrt{N_i}}{\delta_i^2\varepsilon^2 V}. \tag{5.45}$$

Using our definition of δ_i in (5.35), we can upper bound $\sum_{i=m+1}^{I-1} \frac{\sqrt{N_i}}{\delta_i^2}$ by

$$\begin{aligned}\sum_{i=m+1}^{I-1} \frac{\sqrt{N_i}}{\delta_i^2} &\leq \frac{16\zeta(2)^2}{\delta^2} 2^{I/2}\sqrt{\underline{N}} \sum_{i=1}^{I-1} 2^{(i-I)/2}(I-i)^2 \\ &\leq \frac{16\zeta(2)^2}{\delta^2} 2^{I/2}\sqrt{\underline{N}} \sum_{k=1}^{\infty} k^2 2^{-k/2} \leq C2^{I/2}\sqrt{\underline{N}} = C\sqrt{\overline{N}}.\end{aligned} \tag{5.46}$$

Hence the second sum in (5.45) is at most

$$\sum_{i=m+1}^{I-1} C\frac{\sqrt{N_i}}{\delta_i^2\varepsilon^2 V} \leq C\frac{\sqrt{\overline{N}}}{\varepsilon^2 V}. \tag{5.47}$$

We want the right hand side of (5.47) to be $o(1)$, which forces

$$\overline{N} = N_I \ll \varepsilon^4 V^2. \tag{5.48}$$

The bound in (5.48) holds, since $\overline{N} = o(\varepsilon V)$, by the first constraint in (5.38), and since $\varepsilon^3 V \geq 1$.

On the other hand, the first sum in (5.45) can be upper bounded by

$$\sum_{i=0}^{m} \frac{C}{\sqrt{N_i}} \frac{\log n}{n\delta_i^2\varepsilon^2} \leq \frac{C\log n}{n\sqrt{\underline{N}}\varepsilon^2\delta^2} = o(1), \tag{5.49}$$

since $\underline{N} \gg \frac{(\log n)^2}{n^2\varepsilon^4}$ by the third bound in (5.37). This proves the required concentration bound, thus establishing Proposition 2.2 and Theorem 1.1. $\square$

Acknowledgement. The work of RvdH was supported in part by Netherlands Organisation for Scientific Research (NWO). The work of MJL was partly supported by the Nuffield Foundation.

References

[1] N. Alon and J. Spencer. The Probabilistic Method, 2nd Edition. John Wiley and Sons, New York (2000).

[2] K.B. Athreya and P.E. Ney, Branching Processes, Springer, Berlin, 1972.

[3] A.D. Barbour, L. Holst, S. Janson, Poisson Approximation, OUP, Oxford, 1992.

[4] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer, Random subgraphs of finite graphs: I. The scaling window under the triangle condition. *Random Struct. Alg.* **27** (2005), 137–184.

[5] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer, Random subgraphs of finite graphs: II.The lace expansion and the triangle condition. *Ann. Probab.* **33** (2005), 1886–1944.

[6] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer, Random subgraphs of finite graphs: III. The phase transition on the n-cube. *Combinatorica* **26** (2006), 395–410.

[7] C. Borgs, J. T. Chayes, H. Kesten and J. Spencer. Uniform boundedness of critical crossing probabilities implies hyperscaling. *Random Struct. Alg.*, **15** (1999), 368–413.

[8] C. Borgs, J. T. Chayes, H. Kesten and J. Spencer. The birth of the infinite cluster: finite-size scaling in percolation. *Commun. Math. Phys.*, **224** (2001), 153–204.

[9] L.S. Chandran and C.R. Subramanian, A spectral lower bound for the treewidth of a graph and its consequences, preprint, available at www.mpi-sb.mpg.de/∼sunil/applyspectree.ps

[10] L. Devroye, Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms, in Probabilistic Methods for Algorithmic Discrete Mathematics, ed. M.Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed, 249–314, Springer-Verlag, Berlin, 1998.
[11] M. Dwass, The total progeny in a branching process, *J. Appl. Probab.* **6** (1969), 682–686.
[12] M. Heydenreich and R. van der Hofstad. Random graph asymptotics on high-dimensional tori. *Commun. Math. Phys.*, **270** (2007), 335–358.
[13] R. van der Hofstad and G. Slade. Expansion in n^{-1} for percolation critical values on the n-cube and $\mathbb{Z}^n$: the first three terms. *Combin. Probab. Comput.* **15** (2006), 695–713.
[14] R. van der Hofstad and G. Slade. Asymptotic expansions in n^{-1} for percolation critical values on the n-cube and $\mathbb{Z}^n$, *Random Struct. Alg.* **27** (2005), 331–357.
[15] S. Janson, Cycles and unicyclic components in random graphs, *Combin. Probab. Comput.* **12** (2003), 27–52.
[16] S. Janson, On concentration of probability, Contemporary Combinatorics, ed. B. Bollobás, Bolyai Soc. Math. Stud. **10** (2002), János Bolyai Mathematical Society, Budapest, 289–301.
[17] S. Janson, D.E. Knuth, T. Łuczak & B. Pittel, The birth of the giant component, *Random Struct. Alg.* **3** (1993), 233–358.
[18] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs*, Wiley, New York, 2000.
[19] V.F. Kolchin, Moments of degeneration of a branching process and height of a random tree, Mathematical Notes of the Academy of Sciences of the USSR **6** (1978), 954–961.
[20] M.J. Luczak, C. McDiarmid and E. Upfal, On-line routing of random calls in networks, *Probab. Theor. Relat. Fields* **125** (2003), 457–482.
[21] C. McDiarmid, Concentration, in Probabilistic Methods for Algorithmic Discrete Mathematics, ed. M.Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed, 195–248, Springer-Verlag, Berlin, 1998.
[22] A. Nachmias, Mean-field conditions for percolation on finite graphs, preprint.
[23] R. Otter, The multiplicative process, *Ann. Math. Statist.* **20** (1949), 206–224.

Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands.
E-mail address: `rhofstad@win.tue.nl`
URL: `http://www.win.tue.nl/~rhofstad`

Department of Mathematics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom
E-mail address: `m.j.luczak@lse.ac.uk`
URL: `http://www.lse.ac.uk/people/m.j.luczak@lse.ac.uk/`